

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS AGRÓNOMOS

DEPARTAMENTO DE PROYECTOS E INGENIERÍA RURAL

**ANÁLISIS REGIONAL DE LAS
PRECIPITACIONES DIARIAS EXTREMAS
EN LA CUENCA DEL RÍO ARGÁ CON R**

AUTOR

Juan Erro Eslava

TUTOR

José Javier López Rodríguez

INGENIERO AGRÓNOMO

Pamplona, Febrero 2011

ANÁLISIS REGIONAL DE LAS PRECIPITACIONES DIARIAS EXTREMAS EN LA CUENCA DEL RÍO ARGÁ CON R

Tradicionalmente, los métodos de análisis de frecuencia de los eventos extremos correspondientes a las precipitaciones máximas diarias han estado basados en el uso de las series de observaciones a nivel local. Sin embargo, dada la escasa longitud de dichas series y la incertidumbre relacionada con la calidad de las muestras - que presentan por lo general una gran variabilidad del sesgo muestral - este empleo de series locales en la estimación de los cuantiles carece, en cierto modo, de rigor. El método del índice de avenida basado en el estudio de los *L*-momentos, propuesto por Hosking y Wallis (1997), permite resolver esta problemática paliando la carencia de datos en el tiempo con su abundancia en el espacio.

El río Argá, por otra parte, es uno de los ríos más caudalosos de Navarra; atraviesa algunas de las ciudades más importantes del territorio y provoca con cierta frecuencia desbordamientos que se traducen en inundaciones de los núcleos urbanos. Con el fin de caracterizar las frecuencias de las precipitaciones extremas (cuantiles) en la cuenca del río Argá, se propone realizar un análisis regional de las precipitaciones registradas en 24 h. en las estaciones pluviométricas de dicha cuenca. El análisis incluirá los siguientes pasos:

1. filtrado primario de las observaciones,
2. identificación de regiones homogéneas,
3. selección de la función de distribución que mejor se ajuste a cada región,
4. estimación de los cuantiles asociados a un cierto periodo de retorno.

El estudio concluirá con la comparación del análisis regional con los métodos clásicos de análisis local.

La metodología del índice de avenida y el estudio de los *L*-momentos se implementarán mediante el lenguaje de programación R.

Firma del autor:

Firma del tutor:

Fdo.: Juan Erro Eslava

Fdo.: José Javier López Rodríguez

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

1.- INTRODUCCIÓN.....	1
1.1.- Introducción	3
1.2.- Antecedentes	3
1.2.1.- El método del índice de avenida	4
1.2.2.- Introducción a los <i>L</i> -momentos.....	6
1.3.- Objetivos	8
1.4.- Estructura del trabajo	9
2.- CARACTERIZACIÓN DE LA CUENCA Y LAS SERIES, Y DESCRIPCIÓN DEL MÉTODO DE ANÁLISIS REGIONAL	11
2.1.- La cuenca del Arga.....	13
2.2.- Las series de observaciones.....	16
2.3.- Análisis regional de frecuencias.....	19
2.3.1.- Filtrado primario de las observaciones.....	19
2.3.2.- Identificación de regiones homogéneas	21
2.3.3.- Elección de la distribución de frecuencia.....	26
2.3.4.- Estimación de los cuantiles de frecuencia.....	35
2.4.- El entorno de programación R	41
3.- RESULTADOS Y DISCUSIÓN.....	43
3.1.- Análisis regional de frecuencias.....	45
3.1.1.- Filtrado primario de las observaciones.....	45
3.1.2.- Identificación de regiones homogéneas	48
3.1.3.- Elección de la distribución de frecuencia.....	53
3.1.4.- Estimación de los cuantiles de frecuencia.....	65
3.2.- Comparación con los métodos de análisis local.....	88
4.- CONCLUSIONES.....	109
5.- REFERENCIAS	113
ANEXO I: DISTRIBUCIONES Y SUS L-MOMENTOS	121
ANEXO II: IMPLEMENTACIÓN EN R	135

ÍNDICE DE TABLAS

Tabla 2.1.	<i>Datos originales de las estaciones seleccionadas.</i>	18
Tabla 2.2.	<i>Valores críticos de la discordancia D_i.</i>	20
Tabla 2.3.	<i>Pesos de las colas superiores de las distribuciones más comunes.</i>	29
Tabla 3.1.	<i>Valores de los ratios de L-momentos de las estaciones.</i>	45
Tabla 3.2.	<i>Valores de la discordancia D_i de las estaciones.</i>	46
Tabla 3.3.	<i>Descriptores físicos de las estaciones.</i>	48
Tabla 3.4.	<i>Transformación de las características físicas.</i>	49
Tabla 3.5.	<i>Resultado del análisis cluster y medida de la heterogeneidad H.</i>	50
Tabla 3.6.	<i>Resultados ajustados de la medida de heterogeneidad H.</i>	51
Tabla 3.7.	<i>Valores de la discordancia D_i de las estaciones en su región.</i>	51
Tabla 3.8.	<i>Matriz de correlaciones de la región Norte.</i>	53
Tabla 3.9.	<i>Matriz de correlaciones de la región Sur.</i>	54
Tabla 3.10.	<i>Resultados de la medida de heterogeneidad H considerando la dependencia entre estaciones.</i>	54
Tabla 3.11.	<i>Valores del estadístico Z para las diferentes distribuciones en la región Norte.</i>	58
Tabla 3.12.	<i>Error cuadrático medio relativo de las distribuciones candidatas en la región Norte.</i>	59
Tabla 3.13.	<i>Valores del estadístico Z para las diferentes distribuciones en la región Sur.</i>	62
Tabla 3.14.	<i>Error cuadrático medio relativo de las distribuciones candidatas en la región Sur.</i>	63
Tabla 3.15.	<i>Parámetros de la curva regional de frecuencia de la distribución Pearson tipo III de la región Norte.</i>	65
Tabla 3.16.	<i>Resultados de las estimaciones regionales de la región Norte.</i>	66
Tabla 3.17.	<i>Índice de avenida de las estaciones de la región Norte.</i>	67
Tabla 3.18.	<i>Cuantiles de precipitación en 24 h. de Belzunce.</i>	68
Tabla 3.19.	<i>Cuantiles de precipitación en 24 h. de Erro.</i>	69
Tabla 3.20.	<i>Cuantiles de precipitación en 24 h. de Espinal.</i>	70
Tabla 3.21.	<i>Cuantiles de precipitación en 24 h. de Eugui.</i>	71
Tabla 3.22.	<i>Cuantiles de precipitación en 24 h. de Iraizoz.</i>	72
Tabla 3.23.	<i>Cuantiles de precipitación en 24 h. de Iroz.</i>	73
Tabla 3.24.	<i>Cuantiles de precipitación en 24 h. de Olagüe.</i>	74
Tabla 3.25.	<i>Cuantiles de precipitación en 24 h. de Velate.</i>	75
Tabla 3.26.	<i>Cuantiles de precipitación en 24 h. de Zubiri.</i>	76
Tabla 3.27.	<i>Cuantiles de precipitación en 24 h. de las estaciones de la región Norte.</i>	77
Tabla 3.28.	<i>Parámetros de la curva regional de frecuencia de la distribución Pearson tipo III de la región Sur.</i>	78
Tabla 3.29.	<i>Resultados de las estimaciones regionales de la región Sur.</i>	79
Tabla 3.30.	<i>Índice de avenida de las estaciones de la región Sur.</i>	80
Tabla 3.31.	<i>Cuantiles de precipitación en 24 h. de Ilundáin.</i>	81
Tabla 3.32.	<i>Cuantiles de precipitación en 24 h. de Monreal.</i>	82
Tabla 3.33.	<i>Cuantiles de precipitación en 24 h. de Noáin.</i>	83
Tabla 3.34.	<i>Cuantiles de precipitación en 24 h. de Otazu.</i>	84
Tabla 3.35.	<i>Cuantiles de precipitación en 24 h. de Pamplona.</i>	85
Tabla 3.36.	<i>Cuantiles de precipitación en 24 h. de Zuazu.</i>	86
Tabla 3.37.	<i>Cuantiles de precipitación en 24 h. de las estaciones de la región Sur.</i>	87
Tabla 3.38.	<i>Cuantiles de precipitación en 24 h. de Belzunce por el método local.</i>	89
Tabla 3.39.	<i>Cuantiles de precipitación en 24 h. de Erro por el método local.</i>	90
Tabla 3.40.	<i>Cuantiles de precipitación en 24 h. de Espinal por el método local.</i>	91
Tabla 3.41.	<i>Cuantiles de precipitación en 24 h. de Eugui por el método local.</i>	92
Tabla 3.42.	<i>Cuantiles de precipitación en 24 h. de Iraizoz por el método local.</i>	93
Tabla 3.43.	<i>Cuantiles de precipitación en 24 h. de Iroz por el método local.</i>	94
Tabla 3.44.	<i>Cuantiles de precipitación en 24 h. de Olagüe por el método local.</i>	95
Tabla 3.45.	<i>Cuantiles de precipitación en 24 h. de Velate por el método local.</i>	96
Tabla 3.46.	<i>Cuantiles de precipitación en 24 h. de Zubiri por el método local.</i>	97
Tabla 3.47.	<i>Cuantiles de precipitación en 24 h. de Ilundáin por el método local.</i>	100
Tabla 3.48.	<i>Cuantiles de precipitación en 24 h. de Monreal por el método local.</i>	101
Tabla 3.49.	<i>Cuantiles de precipitación en 24 h. de Noáin por el método local.</i>	102
Tabla 3.50.	<i>Cuantiles de precipitación en 24 h. de Otazu por el método local.</i>	103
Tabla 3.51.	<i>Cuantiles de precipitación en 24 h. de Pamplona por el método local.</i>	104
Tabla 3.52.	<i>Cuantiles de precipitación en 24 h. de Zuazu por el método local.</i>	105

ÍNDICE DE FIGURAS

Fig. 2.1.	<i>Situación de la cuenca del Arga</i>	13
Fig. 2.2.	<i>Zona de la cuenca acotada para el análisis regional de precipitaciones</i>	15
Fig. 2.3.	<i>Estaciones meteorológicas manuales de Navarra</i>	16
Fig. 2.4.	<i>Estaciones empleadas en el análisis regional</i>	17
Fig. 2.5.	<i>Representación gráfica del concepto de discordancia</i>	19
Fig. 2.6.	<i>Comparativa entre los datos observados y los datos simulados</i>	24
Fig. 2.7.	<i>Representación gráfica del concepto de bondad de ajuste</i>	32
Fig. 3.1.	<i>Representación gráfica de los ratios de las estaciones</i>	46
Fig. 3.2.	<i>Representación gráfica de Iroz en la región Sur</i>	50
Fig. 3.3.	<i>Diagrama de L-momentos de las distribuciones en la región Norte</i>	57
Fig. 3.4.	<i>Curva regional de frecuencia de las distribuciones candidatas de la región Norte</i>	58
Fig. 3.5.	<i>Diagrama de L-momentos de las distribuciones en la región Sur</i>	61
Fig. 3.6.	<i>Curva regional de frecuencia de las distribuciones candidatas de la región Sur</i>	62
Fig. 3.7.	<i>Representación gráfica de las estimaciones regionales de la región Norte</i>	66
Fig. 3.8.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Belzunce</i>	68
Fig. 3.9.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Erro</i>	69
Fig. 3.10.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Espinal</i>	70
Fig. 3.11.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Eugui</i>	71
Fig. 3.12.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Iraizoz</i>	72
Fig. 3.13.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Iroz</i>	73
Fig. 3.14.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Olagüe</i>	74
Fig. 3.15.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Velate</i>	75
Fig. 3.16.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Zubiri</i>	76
Fig. 3.17.	<i>Representación gráfica de las estimaciones regionales de la región Sur</i>	79
Fig. 3.18.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Ilundáin</i>	81
Fig. 3.19.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Monreal</i>	82
Fig. 3.20.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Noáin</i>	83
Fig. 3.21.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Otazu</i>	84
Fig. 3.22.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Pamplona</i>	85
Fig. 3.23.	<i>Representación gráfica de los cuantiles de precipitación en 24 h. de Zuazu</i>	86
Fig. 3.24.	<i>Representación de los cuantiles de precipitación en 24 h. de Belzunce por el método local</i>	89
Fig. 3.25.	<i>Representación de los cuantiles de precipitación en 24 h. de Erro por el método local</i>	90
Fig. 3.26.	<i>Representación de los cuantiles de precipitación en 24 h. de Espinal por el método local</i>	91
Fig. 3.27.	<i>Representación de los cuantiles de precipitación en 24 h. de Eugui por el método local</i>	92
Fig. 3.28.	<i>Representación de los cuantiles de precipitación en 24 h. de Iraizoz por el método local</i>	93
Fig. 3.29.	<i>Representación de los cuantiles de precipitación en 24 h. de Iroz por el método local</i>	94
Fig. 3.30.	<i>Representación de los cuantiles de precipitación en 24 h. de Olagüe por el método local</i>	95
Fig. 3.31.	<i>Representación de los cuantiles de precipitación en 24 h. de Velate por el método local</i>	96
Fig. 3.32.	<i>Representación de los cuantiles de precipitación en 24 h. de Zubiri por el método local</i>	97
Fig. 3.33.	<i>Representación de los cuantiles de precipitación en 24 h. de Ilundáin por el método local</i>	100
Fig. 3.34.	<i>Representación de los cuantiles de precipitación en 24 h. de Monreal por el método local</i>	101
Fig. 3.35.	<i>Representación de los cuantiles de precipitación en 24 h. de Noáin por el método local</i>	102
Fig. 3.36.	<i>Representación de los cuantiles de precipitación en 24 h. de Otazu por el método local</i>	103
Fig. 3.37.	<i>Representación de los cuantiles de precipitación en 24 h. de Pamplona por el método local</i>	104
Fig. 3.38.	<i>Representación de los cuantiles de precipitación en 24 h. de Zuazu por el método local</i>	105

1.- INTRODUCCIÓN

1.1.- Introducción

1.2.- Antecedentes

1.2.1.- El método del índice de avenida

1.2.2.- Introducción a los L -momentos

1.3.- Objetivos

1.4.- Estructura del trabajo

1.1.- INTRODUCCIÓN

Junto a sus afluentes Ulzama y Elorz, el río Arga ha sido fuente de asentamientos a lo largo de la historia, convirtiendo a Pamplona y su Comarca en el núcleo de población principal de Navarra. En las últimas décadas, la reducida superficie del término municipal de Pamplona y la escasez de espacios libres en el mismo han hecho que los diferentes desarrollos urbanísticos hayan ejercido una cierta presión sobre estas áreas pertenecientes al cauce principal de los ríos. De esta forma, con cierta frecuencia se producen avenidas que causan inundaciones en zonas urbanizadas residenciales con los consiguientes daños sobre viviendas, mobiliario, etc.

Ante esta situación y como paso previo a la simulación hidráulica que permita simular los niveles de la lámina de agua en las secciones transversales del río y, por lo tanto, las manchas o zonas inundables, es importante caracterizar hidrológicamente la cuenca vertiente mediante el estudio de su precipitación.

1.2.- ANTECEDENTES

El objetivo fundamental del análisis de frecuencias es la estimación de los sucesos extremos (precipitaciones máximas, avenidas, etc.) correspondientes a diferentes períodos de retorno mediante el uso de funciones de distribución de probabilidad. Cabe reseñar que esta estimación de la frecuencia con que los eventos extremos acontecen es compleja, dado que son por definición excepcionales y los registros de las series de observaciones son a menudo demasiado cortos para poder extrapolar con criterio.

Tradicionalmente, la mayoría de los métodos de análisis utilizados han estado basados en el uso local de las series de observaciones. Valga como ejemplo la gran aceptación de la función biparamétrica de distribución de valores extremos de Gumbel: su uso forma parte de la actividad cotidiana en el análisis de frecuencia extrema de las precipitaciones máximas a nivel de estación puntual, habiendo arrojado aceptables resultados aplicados en diferentes ámbitos de la ingeniería (MOPTMA, 1995; Gómez *et al.*, 1996; MMA, 1998). Sin embargo, el uso de datos locales en la estimación de los cuantiles de frecuencia de los eventos extremos puede no justificarse siempre con rigor, debido fundamentalmente a la escasa longitud de las series de observaciones y a la incertidumbre relacionada con la calidad de las muestras, que presentan por lo general una gran variabilidad del sesgo muestral.

De acuerdo con Hosking y Wallis (1997), este problema puede resolverse “*paliando la carencia de datos en el tiempo con su abundancia en el espacio*”, pues son muchas las situaciones en las que se dispone de observaciones en diferentes estaciones cercanas entre sí. En esta idea se cimienta el análisis regional de frecuencias; constituye una alternativa útil y toma en consideración las deficiencias de los métodos de análisis local, aumentando la longitud de los registros disponibles mediante la transferencia de información entre las diferentes estaciones que componen una supuesta región. A modo de ejemplo, si se cuenta con N sitios, cada uno de ellos con n años de registro de eventos máximos, puede esperarse que $N \times n$ datos darán estimaciones más precisas de cuantiles tan extremos como Q_{Nn} .

El análisis regional de frecuencias emplea el método del índice de avenida conjuntamente con el uso de los momentos lineales (*L*-momentos). Se trata de una metodología robusta aun en presencia de datos anómalos (*outliers*), y su buen uso puede conducir a estimaciones suficientemente precisas y de gran utilidad en diversos problemas reales. Si bien su empleo a nivel internacional viene siendo habitual, con estudios importantes en Australia (Pearson *et al.*, 1991), Estados Unidos (Vogel *et al.*, 1993), Nueva Zelanda (Madsen *et al.*, 1997), Italia (Norbiato *et al.*, 2007) y China (Yang *et al.*, 2010), entre otros, en España apenas se ha utilizado. Las investigaciones del Centro de Estudios Hidrográficos del CEDEX para caracterizar los caudales de avenida de los ríos peninsulares (Jiménez y Mediero, 2009) o los trabajos de Álvarez *et al.* (1999) para las precipitaciones de Galicia, y García (2007) para Andalucía constituyen algunos de los ejemplos.

1.2.1.- EL MÉTODO DEL ÍNDICE DE AVENIDA

El término “índice de avenida”, aplicado por vez primera en el estudio de las avenidas por Dalrymple (1960) y de ahí su nombre, se ha utilizado desde entonces en el análisis de frecuencia de otras variables diferentes a la original, entre ellas las precipitaciones máximas (p.e. Saénz de Ormijana *et al.*, 1991; Guttman, 1993; Ferrer y Ardiles, 1994). El principio subyacente del método es que la distribución de los eventos máximos en una *región homogénea* es la misma excepto por un factor de escala (el índice de avenida) que particulariza las características de dichos eventos en cada estación. Conocidos los cuantiles a escala regional y los valores medios locales, se pueden extrapolar extremos locales con cierta fiabilidad.

Supónganse disponibles los datos de una región con N estaciones, cada estación i con un tamaño muestral n_i . Los datos observados en cada estación i son Q_{ij} , donde $j = 1, \dots, n_i$. El principio en el que se basa el método establece que la función cuantil de la distribución de frecuencia de la estación i , es decir, la función inversa de dicha distribución, $Q_i(F)$ $0 < F < 1$, se define como:

$$Q_i(F) = \mu_i q(F), \quad i = 1, \dots, N. \quad (1.1)$$

donde μ_i es el índice de avenida, que toma el valor de la media de la distribución de frecuencia en la estación i considerada, y cuyo estimador $\hat{\mu}_i$ es la media de las observaciones en dicha estación; y $q(F)$ es la curva regional de frecuencia, una función cuantil adimensional común a todas las estaciones de la región.

Esta curva regional, $q(F)$, es función de un conjunto, p , de parámetros a establecer, $\hat{\theta}_k$ con $k = 1, \dots, p$, a los que se les puede asignar los momentos de grado superior de la función, como el coeficiente de variación, el de sesgo, o incluso determinados cocientes de *L*-momentos. Por lo tanto, la curva regional también se puede expresar como $q(F, \hat{\theta}_1, \dots, \hat{\theta}_p)$. La base para estimar la curva regional de frecuencia la constituyen las observaciones adimensionalizadas $q_{ij} = Q_{ij} / \hat{\mu}_i$, siendo $j = 1, \dots, n_i$, e $i = 1, \dots, N$. Una vez estimados dichos parámetros en cada estación por separado, se combinan para dar

una estimación regional de cada parámetro, $\hat{\theta}_k^R$, a modo de media ponderada, de manera que cada estación, i , tiene un peso proporcional al tamaño de su muestra, n_i , es decir:

$$\hat{\theta}_k^R = \frac{\sum_{i=1}^N n_i \hat{\theta}_k^{(i)}}{\sum_{i=1}^N n_i}. \quad (1.2)$$

Sustituyendo los p parámetros regionales estimados se obtiene la estimación de la curva regional de frecuencia $\hat{q}(F; \hat{\theta}_1^R, \dots, \hat{\theta}_p^R)$.

Obtenidos tanto el estimador $\hat{\mu}_i$ del índice de avenida como el estimador $\hat{q}(F)$ de la curva regional de frecuencia, la estimación de los cuantiles de interés en una estación i se reduce a la aplicación de la ecuación de partida:

$$\hat{Q}_i(F) = \hat{\mu}_i \hat{q}(F). \quad (1.3)$$

Conviene señalar que el método del índice de avenida parte de asumir una serie de supuestos:

- i. Las observaciones de una estación están idénticamente distribuidas.
- ii. Las observaciones en una estación son independientes.
- iii. Observaciones en diferentes estaciones son independientes.
- iv. Las distribuciones de frecuencia en distintos sitios de una misma región son idénticas excepto por un factor de escala.
- v. La forma de la curva regional de frecuencias se especifica correctamente desde el punto de vista matemático.

Las dos primeras suposiciones se cumplen para la mayor parte de los casos. Tanto los valores máximos como los acumulados anuales están, por lo general, libres de variaciones temporales. Además, es evidente que todos los métodos de análisis de frecuencia han de asumir que los eventos observados en el pasado constituyen la primera pista sobre lo que se puede esperar en el futuro. Por otro lado, Landwehr *et al.* (1979) y McMahon y Srikanthan (1982) estudiaron el efecto de la dependencia serial en el análisis local, concluyendo que ésta tan sólo provoca un pequeño aumento en el sesgo y el error estándar de los cuantiles estimados.

Sin embargo, la tercera suposición rara vez se cumple estrictamente. En estudios habituales de precipitaciones y caudales, por ejemplo, cabe esperar la existencia de una cierta correlación entre las observaciones de estaciones cercanas entre sí. Más aún: es obvio que las dos últimas suposiciones nunca se van a cumplir en la práctica. Como mucho podrán ser alcanzadas aproximadamente, mediante selección cuidadosa de las estaciones de cada región y elección de distribuciones de frecuencia consistentes con las observaciones.

Así pues, el método del índice de avenida será apropiado siempre y cuando constituya un modelo robusto que aproxime a estos cinco supuestos.

1.2.2.- INTRODUCCIÓN A LOS *L*-MOMENTOS

Los momentos lineales surgen de combinaciones lineales de los *momentos ponderados probabilísticamente*, introducidos por Greenwood *et al.* (1979), y constituyen un sistema alternativo al método tradicional de los momentos convencionales para describir las funciones de distribución. Los momentos ponderados probabilísticamente, de la variable aleatoria X con una función de distribución $F(X)$, quedan definidos por la expresión:

$$M_{p,k,s} = E\left(X^p \{F(X)\}^k \{1 - F(X)\}^s\right) \quad (1.4)$$

donde E indica el valor de la Esperanza, siendo p , k y s números reales.

Son particularmente útiles los casos en que $p = 1$ con $k = 0$, y $p = 1$ con $s = 0$, denotados como $\alpha_r = M_{1,0,r}$ y $\beta_r = M_{1,r,0}$, respectivamente, siendo $r = 0, 1, 2, \dots$

En términos de estos momentos de probabilidad ponderada, los cuatro primeros momentos lineales λ_i quedan definidos por las siguientes expresiones:

$$\begin{aligned} \lambda_1 &= \alpha_0 = \beta_0, \\ \lambda_2 &= \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0, \\ \lambda_3 &= \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0, \\ \lambda_4 &= \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0. \end{aligned}$$

El momento lineal de primer orden λ_1 es el parámetro de localización o media de la distribución; el de segundo orden λ_2 mide la escala de la distribución, indicando el grado de dispersión de los datos; el momento de tercer orden λ_3 hace referencia a su asimetría; y el de cuarto orden λ_4 indica la kurtosis (apuntamiento) de la distribución.

No obstante, para poder comparar diferentes estaciones resulta imprescindible definir los momentos lineales de forma adimensional. Por ello, se calculan los ratios o cocientes de momentos lineales, dividiendo aquellos de mayor orden por la medida de escala λ_2 (Hosking y Wallis, 1997):

$$\tau_r = \lambda_r / \lambda_2, \quad r = 3, 4, \dots \quad (1.5)$$

dando así una idea de la forma de la distribución con independencia de su escala de medida. Adquieren particular importancia los *L*-coeficientes de variación, asimetría y kurtosis:

$$L - CV = \tau = \frac{\lambda_2}{\lambda_1} \quad L - \text{asimetría} = \tau_3 = \frac{\lambda_3}{\lambda_2} \quad L - \text{kurtosis} = \tau_4 = \frac{\lambda_4}{\lambda_2}$$

Se puede afirmar que, conceptualmente, los L -momentos son semejantes a los momentos convencionales. Sin embargo, Hosking y Wallis (1997) constatan que son varias las ventajas de los primeros desde el punto de vista teórico: capacidad para caracterizar un mayor rango de funciones de distribución, mayor robustez en presencia de *outliers* y menor sensibilidad a la asimetría, entre otras.

A nivel práctico, los L -momentos con los que se tiende a trabajar son la media λ_1 , la escala λ_2 , el L -CV τ , el L -Asimetría τ_3 , y el L -kurtosis τ_4 , expresados respectivamente como l_1 , l_2 , t , t_3 y t_4 cuando hacen referencia a los valores muestrales de las observaciones.

1.3.- OBJETIVOS

Con el fin de caracterizar las frecuencias de las precipitaciones extremas en la cuenca del río Arga, se propone:

- Desarrollar un análisis regional de las precipitaciones registradas en 24 h. en las estaciones pluviométricas de dicha cuenca, empleando el método del índice de avenida conjuntamente con el uso de los *L*-momentos.
- Estimar los valores de los cuantiles de precipitación en 24 h. asociados a diferentes periodos de retorno.
- Comparar los resultados obtenidos del análisis regional con los que se conseguirían mediante el análisis clásico local.

A nivel metodológico, también se plantea como objetivo del trabajo desarrollar una herramienta en R - lenguaje y entorno de programación para análisis estadístico y gráfico (*R Development Core Team*, 2010) - que permita realizar este tipo de estudios de forma eficiente y rápida.

1.4.- ESTRUCTURA DEL TRABAJO

El trabajo comenzará con una descripción somera de la cuenca del río Arga y de las series de precipitación en 24 h. empleadas en el estudio.

A continuación se procederá a desarrollar el análisis regional de frecuencias, que consta de cuatro etapas:

1. *Filtrado primario de las observaciones.* Es el primer paso en cualquier tipo de análisis estadístico. Está encaminado a detectar posibles errores o inconsistencias en unos datos de partida que han de ser homogéneos en el tiempo. Por ello es útil cualquier información externa acerca de los métodos de toma de medidas, o sobre cualquier cambio en el tiempo que haya podido afectar a la distribución de frecuencia en alguna de las estaciones de estudio.
2. *Identificación de regiones homogéneas.* Consiste en repartir los diferentes lugares en regiones, dado que la región es la unidad fundamental del análisis regional. Las estaciones no pueden incluirse en regiones en las que el criterio de homogeneidad no se satisfaga, si bien una homogeneidad aproximada puede ser suficiente para garantizar precisión en las estimaciones. Calculando diversos estadísticos en cada lugar y comparando la variabilidad entre estaciones, así como con los valores esperados para la región, puede estudiarse el grado de homogeneidad de la misma. Los L -momentos son buenos estadísticos para realizar la prueba de heterogeneidad.
3. *Elección de una distribución de frecuencia para la región.* Consiste en seleccionar la distribución regional de frecuencia, $q(F)$, más apropiada para la región, mediante pruebas de bondad de ajuste a las diversas funciones posibles. Hosking y Wallis (1997) recomiendan evaluar las distribuciones Logística Generalizada (*Generalized Logistic*, GLO), General de Valores Extremos (*Generalized Extreme Value*, GEV), General de Pareto (*Generalized Pareto*, GPA), LogNormal de tres parámetros (*LogNormal 3*, LN3) y Pearson tipo III (*Pearson type III*, PE3), por incluir entre sus casos particulares las distribuciones más empleadas en el estudio de variables ambientales. Es necesario determinar el grado de similitud existente entre los estadísticos referentes a los datos observados y los que se esperarían si estos procediesen de una muestra aleatoria extraída en una población que obedeciese a una función de distribución de frecuencias conocida.
4. *Estimación de los cuantiles regionales de frecuencia.* Escogida la función de distribución adecuada para la región, se estiman los cuantiles asociados a un cierto periodo de retorno. Para ello se emplea el algoritmo regional de los L -momentos, a través del cual los parámetros que definen la distribución se estiman separadamente en cada estación, combinándose después para obtener una media regional. Con el método del índice de avenida, los valores regionales se particularizan en cada estación local. Además, para que los resultados obtenidos alcancen la mayor utilidad práctica posible y para poder calibrar su fiabilidad, se evalúa la magnitud de la incertidumbre de las estimaciones mediante simulación de tipo Monte Carlo.

Desarrollado el análisis regional, el trabajo concluirá con la valoración de la robustez del modelo por comparación con los métodos tradicionales de análisis local.

Toda la metodología descrita será implementada mediante lenguaje de programación y software R (*R Development Core Team*, 2010), cuyo paquete estadístico permitirá ejecutar diferentes órdenes de cálculo previamente codificadas.

2.- CARACTERIZACIÓN DE LA CUENCA Y LAS SERIES, Y DESCRIPCIÓN DEL MÉTODO DE ANÁLISIS REGIONAL

2.1.- La cuenca del Arga

2.2.- Las series de observaciones

2.3.- Análisis regional de frecuencias

2.3.1.- Filtrado primario de las observaciones

2.3.2.- Identificación de regiones homogéneas

2.3.3.- Elección de la distribución de frecuencia

2.3.4.- Estimación de los cuantiles de frecuencia

2.4.- El entorno de programación R

2.1.- LA CUENCA DEL ARGÁ



Fig. 2.1. Situación de la cuenca del Arga (Fuente: Departamento de Medio Ambiente, Ordenación del Territorio y Vivienda – Gobierno de Navarra, 2007).

La cuenca del río Arga, englobada dentro de la cuenca hidrográfica del Ebro, comprende el territorio de Navarra drenado por el río que le da nombre y varios afluentes. Con una longitud de aproximada de 150 km, el Arga recoge aguas de una cuenca vertiente de unos 2760 km², casi todos ellos pertenecientes a Navarra (2550 km²) a excepción de una pequeña parte de la cuenca del Araquil, que es territorio del País Vasco (ver figura 2.1).

El río Arga nace en la vertiente meridional del collado de Urquiaga (Quinto Real), situada en la divisoria cantábrico-mediterránea, al NO del valle de Erro. Poco después toma dirección sur, siendo sus aguas represadas en el embalse de Eugui que, con una capacidad de 21 hm³, constituye el principal reservorio de agua potable para la

Mancomunidad de Aguas de la Comarca de Pamplona. Una vez atravesado éste, continúa su curso por el valle de Esteribar hasta llegar a la entrada de la cuenca de Pamplona, donde recibe al río Ulzama – afluente de gran importancia que duplica su caudal - en el municipio de Huarte. Hasta este punto, tanto el Arga como el Ulzama se caracterizan por ser ríos de montaña, con pendientes acusadas que discurren por valles estrechos. Al atravesar la ciudad de Pamplona, la pendiente se reduce considerablemente y adquiere un recorrido meandriforme, circulando entre extensas plataformas inundables que en momentos de riadas pueden verse anegadas. Poco después de recoger las aguas del Elorz, al dejar Pamplona, el Arga se une al río Araquil en Ibero para seguir aguas abajo. Desemboca en el río Aragón en las cercanías del municipio de Funes, a poca distancia de la desembocadura de éste en el río Ebro.

Por tanto, el río Arga presenta una dirección dominante norte-sur, y en su recorrido recoge la aportación de un buen número de afluentes. Incrementa considerablemente su caudal en la cuenca de Pamplona, gracias a las aportaciones del río Ulzama y del río Araquil principalmente y, en menor medida, de los ríos Urbi, Sadar y Elorz. Fuera de la cuenca, el Arga recibe aguas del río Robo en Puente La Reina y del río Salado en Mendigorriá. A partir de esta localidad, las aportaciones de otros afluentes son escasas: Lagortea, en Artajona, y el Barranco de la Nava, en Larraga.

Esta disposición condicionada por la topografía permite hablar de tres regiones principales en la cuenca:

- Zona norte, que es la que presenta las mayores altitudes, y en donde los ríos tienen mayores pendientes. La excepción de esta parte de la cuenca es el río Araquil, que al discurrir de oeste a este presenta una pendiente más baja que la del resto de los ríos de cabecera. Esta zona alta es donde más llueve, y los cursos de los ríos se encuentran por encima de los 400 m.
- Zona media, entre la desembocadura del Araquil y Larraga, en la que los cursos de los ríos se encuentran entre los 400 y 300 m. En esta franja las pendientes de los ríos son menores al ser el relieve menos abrupto que en cabecera.
- Zona sur, a partir de Larraga, en la que los cursos de los ríos se sitúan por debajo de los 300 m, y caracterizada por un relieve mucho más llano.

En el Documento Técnico para la Participación Pública en la Cuenca del Arga (Foro del Agua de Navarra – Gobierno de Navarra, 2007) se sintetizan las características climáticas, geográficas e hidrológicas más relevantes de la cuenca. Se establece que la aportación media del río Arga es de 1559 hm³/año, suponiendo el 15,5% del total de los recursos de Navarra, si bien una pequeña parte de ellos (81 hm³/año) se generan fuera de la Comunidad.

El Documento define una precipitación media de la cuenca para el periodo 1940-2000 de 972 mm/año, variando entre los 1600 mm/año en zonas altas y los 450 mm/año en la desembocadura. Las precipitaciones más abundantes se producen entre noviembre y enero, mientras que el mínimo pluviométrico se registra en los meses de julio y agosto. La evapotranspiración (ETP) media adopta valores en torno a 700 mm/año, siendo inferior a 650 en algunas zonas de las cabeceras y mayor de 750 en la desembocadura.

Si se compara el valor medio de la ETP con el de la precipitación en la cuenca, se pone de relieve que los meses más secos son los que van de mayo a septiembre. En lo referente a la temperatura media anual, varía entre 10°C en las zonas montañosas del norte de la cuenca y 14°C en la zona más baja, con temperaturas más altas en los meses de julio y agosto, y menores de diciembre a febrero. Se puede afirmar, por tanto, que la cuenca del Arga, al igual que Navarra, se caracteriza por una transición climática y biogeográfica muy fuerte, con un cambio paulatino desde el clima de tipo oceánico-montañoso en cabecera, influido por la altitud, hasta un clima de tipo mediterráneo-continental en la zona de desembocadura.

Por otro lado, en el contexto del análisis regional de frecuencias de las precipitaciones máximas, conviene acotar la zona de estudio. Y es que, pese a que el río Arga cruza buena parte del territorio navarro, la mayor importancia socio-económica y demográfica recae en la cuenca de Pamplona, donde sus avenidas pueden causar y causan cuantiosos daños en infraestructuras, edificios y zonas agrícolas, tal y como ha ocurrido a lo largo de la historia con relativa frecuencia.

Así, la figura 2.2 refleja la zona acotada para el análisis. En ella se representa la cuenca del río Arga que vierte al punto en el que se ubica la estación de aforo de Arazuri, perteneciente a la Confederación Hidrográfica del Ebro. Esta estación se localiza poco después de la confluencia del Arga con el río Elorz, cuando el Arga abandona Pamplona. La superficie de esta cuenca es de 803,49 km², y la red fluvial queda definida fundamentalmente por tres ríos y sus respectivas subcuencas: Arga, Ulzama y Elorz. Es ahí donde se centra el estudio.

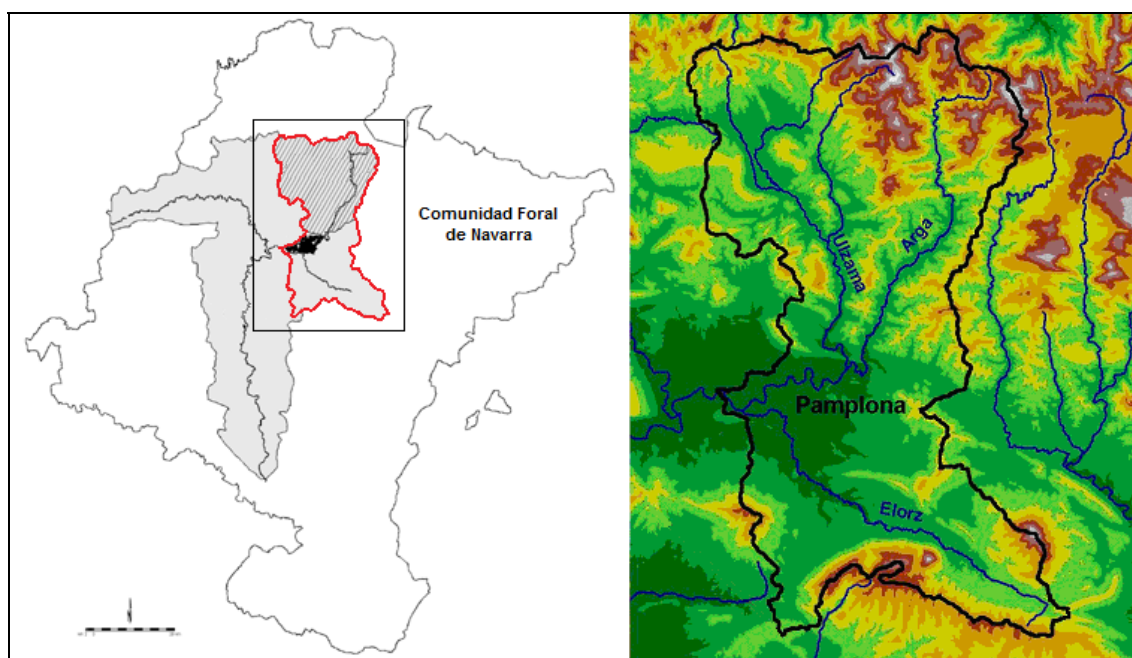


Fig. 2.2. Zona de la cuenca acotada para el análisis regional de precipitaciones.

2.2.- LAS SERIES DE OBSERVACIONES

Las observaciones empleadas en el análisis corresponden a los valores de precipitación en 24 h. registrados en parte de las estaciones meteorológicas manuales que se distribuyen por Navarra.

Accesibles a través del portal de Meteorología y Climatología de Navarra (<http://meteo.navarra.es/>) - coordinado por los Departamentos de Desarrollo Rural e Innovación del Gobierno de Navarra - se ordenan en archivos de formato “.txt”. Cada archivo corresponde a un año, y en él se recogen los datos de precipitación de los 365 días de que consta. Así, existen tantos archivos por estación como años de registro.

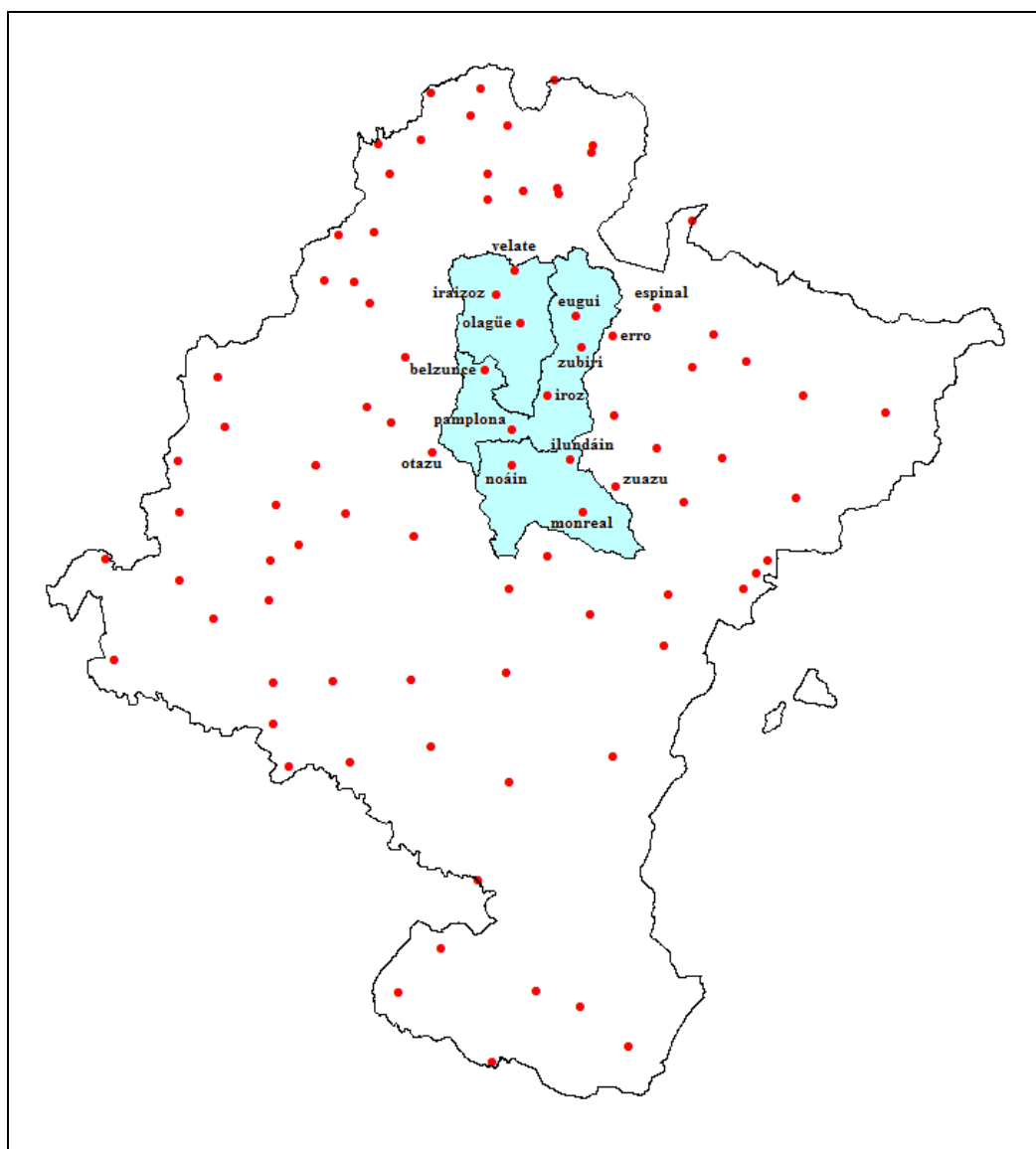


Fig. 2.3. *Estaciones meteorológicas manuales de Navarra* (Fuente: IDENA, Infraestructura de Datos Espaciales de Navarra, 2006).

En la figura 2.3 se representan en puntos rojos las 88 estaciones meteorológicas manuales existentes en Navarra, nombrándose únicamente las 15 estaciones de la región acotada de la cuenca del Arga y sus alrededores que se consideran en el análisis: Belzunce, Erro, Espinal, Eugui, Ilundáin, Iraizoz, Iroz, Monreal, Noáin, Olagüe, Otazu, Pamplona, Velate, Zuazu y Zubiri.

La justificación de la elección de estas 15 estaciones radica tanto en la delimitación de la región de estudio como en los mismos fundamentos del análisis regional de frecuencias. Por un lado, es evidente la necesidad de abarcar adecuadamente la zona en la que se quieren caracterizar las precipitaciones máximas. Pero, por otra parte, la idea en que se cimienta el método regional de paliar la carencia de datos en el tiempo con su abundancia en el espacio no tiene por qué implicar el empleo de todas las estaciones disponibles. Es la integridad física quien marca su requerimiento. Así, se consideran las estaciones de Erro, Espinal, Otazu y Zuazu porque, pese a no quedar incluidas en la cuenca, se aprecian como representativas y coherentes con la región. No ocurre lo mismo con Aoiz, Olóriz, Zalba y las estaciones pertenecientes a la cuenca del Araquil, donde el régimen de precipitaciones y la topografía son tan diferentes que pueden incluso inducir a una estimación menos precisa de los cuantiles. La figura 2.4 aporta una mejor visión de esta explicación.

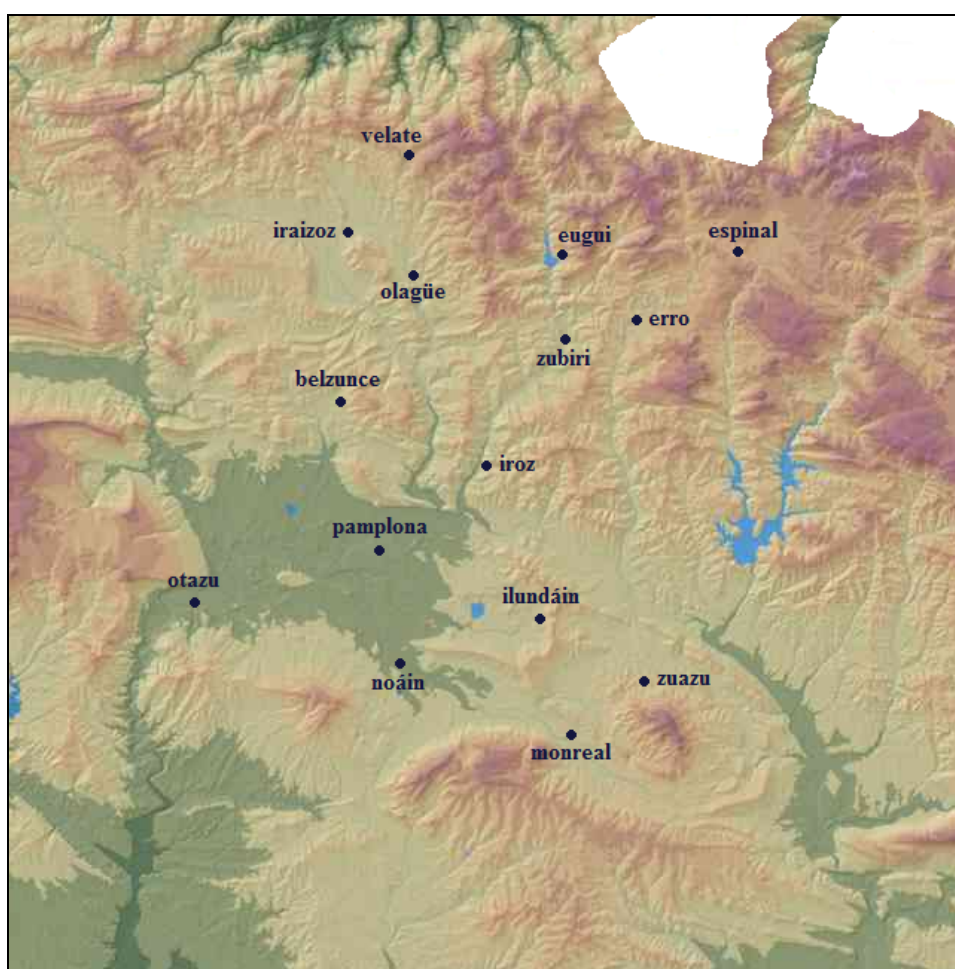


Fig. 2.4. Estaciones empleadas en el análisis regional (Fuente: SITNA, Sistema de Información Territorial de Navarra).

En ella se refleja cómo las estaciones consideradas cubren buena parte de la comarca de Pamplona y la cabecera de la cuenca del Arga, distinguiéndose claramente la zona llana de la montañosa.

Así pues, se presenta el cuadro resumen de las estaciones seleccionadas para el análisis regional de precipitaciones máximas, indicándose número y rango de años de registro de cada una de ellas:

Tabla 2.1. *Datos originales de las estaciones seleccionadas.*

Estación	Longitud (años)	Rango de años
Belzunce	21	1988-2008
Erro	34	1975-2008
Espinal	37	1972-2008
Eugui	41	1968-2008
Ilundáin	30	1979-2008
Iraizoz	39	1970-2008
Iroz	34	1975-2008
Monreal	29	1929-1931; 1983-2008
Noáin	34	1975-2008
Olagüe	34	1975-2008
Otazu	33	1976-2008
Pamplona	123	1880-1919; 1921-1948; 1953-1975; 1977-2008
Velate	22	1987-2008
Zuazu	22	1987-2008
Zubiri	41	1929-1933; 1966-1967; 1975-2008

2.3.- ANÁLISIS REGIONAL DE FRECUENCIAS

2.3.1.- FILTRADO PRIMARIO DE LAS OBSERVACIONES

Comprobar si los datos de los que se parte son apropiados es el primer y esencial paso para cualquier tipo de análisis de frecuencias. Las observaciones recogidas en cada estación han de constituir una muestra homogénea y representativa de la variable que se va a estudiar.

Para el caso concreto de las precipitaciones, dos tipos de errores son habituales en las muestras: por un lado, posibles fallos en el grabado o la transcripción de los datos, que conducen a valores incorrectos; por otro, posibles cambios en las circunstancias bajo las cuales la variable fue medida, como puede ser el reemplazo de los aparatos de medición o su traslado de un lugar a otro. En caso de existir, ambos errores conducen a una distribución de frecuencias no constante en el tiempo. De ahí que neutralizarlos se convierta en labor imprescindible.

Una revisión individualizada de las muestras o la simple comparación entre valores de estaciones cercanas permiten realizar un primer filtrado. Los tests de detección de *outliers* y tendencias, consolidados en la literatura estadística (Kendall, 1975; Barnett y Lewis, 1994), constituyen una alternativa interesante, así como las técnicas de doble masa o los diagramas cuantil-cuantil, muy empleados en el estudio de series temporales. Sin embargo, en el contexto del análisis regional de frecuencias, Hosking y Wallis (1997) establecen que los cambios de tendencia, *outliers*, valores incorrectos y demás errores de una muestra quedan reflejados en sus L -momentos. Definen así la medida de la discordancia, un estadístico que compara los L -momentos de cada estación en una región con los del resto de estaciones y los valores medios de dicha región.

Así, se puede decir que, dado un grupo de estaciones, el objetivo de la discordancia es determinar aquéllas que son discordantes con el grupo. Definiendo el vector de L -momentos (L -CV, L -asimetría, L -kurtosis) de cada estación y considerándolo como un punto en el espacio tridimensional, el grupo de estaciones da lugar a una nube de puntos. Se considera discordante cualquier punto que quede “lejos” del centro de la nube.

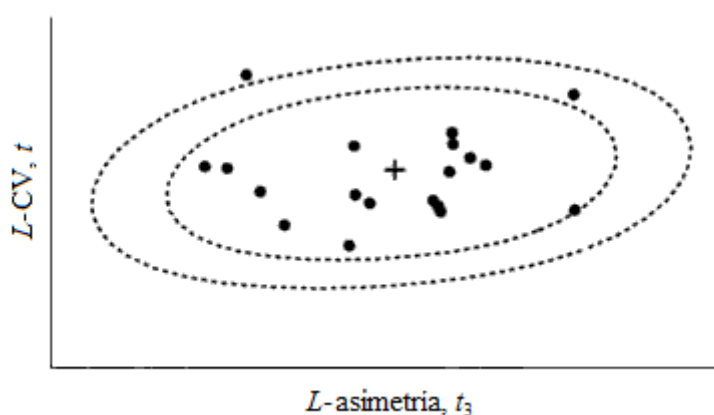


Fig. 2.5. Representación gráfica del concepto de discordancia.

Sirva como ejemplo gráfico la figura 2.5, donde por conveniencia se han representado únicamente los valores del L -CV y L -asimetría. El centro de la nube de puntos, marcado como +, corresponde a la media del grupo. Se construyen elipses concéntricas cuyos ejes mayor y menor son definidos para dar el mejor ajuste posible a los datos, de forma que son discordantes los puntos que quedan fuera de su alcance.

Numéricamente, supónganse N estaciones en el grupo. Sea $u_i = [t^{(i)} \quad t_3^{(i)} \quad t_4^{(i)}]^T$ el vector que contiene los valores de los ratios de L -momentos t , t_3 y t_4 de la estación i , con el superíndice T indicando transposición del mismo. Sea

$$\bar{u} = N^{-1} \sum_{i=1}^N u_i \quad (2.1)$$

la media del grupo (no ponderada). Se define la matriz de suma de cuadrados y productos cruzados,

$$A = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T. \quad (2.2)$$

Por último, se define la medida de la discordancia del sitio i ,

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}). \quad (2.3)$$

La estación i se considera discordante si el valor de D_i es elevado, y ello está condicionado por el número de estaciones en el grupo. Hosking y Wallis (1997) establecen que una estación ha de ser catalogada como discordante si su D_i supera un valor crítico establecido en función del número de sitios en la región:

Tabla 2.2. Valores críticos de la discordancia D_i .

Número de estaciones en la región	Valor crítico	Número de estaciones en la región	Valor crítico
5	1.333	10	2.491
6	1.648	11	2.632
7	1.917	12	2.757
8	2.140	13	2.869
9	2.329	14	2.971
		≥ 15	3

A nivel práctico, se recomienda el empleo de la medida de discordancia en dos ocasiones. En primer lugar, al comienzo del análisis, procurando abarcar la totalidad de las estaciones o regiones suficientemente amplias como para que su número sea elevado. Con esta primera aproximación se trata de detectar aquellos sitios que difieren significativamente de la mayoría, sea por la presencia de valores atípicos o por la

existencia de errores de bulto. Evidentemente, requerirán de una revisión más exhaustiva y profunda para constatar si ha habido fallos en la ejecución de la medida o en su transcripción, y en su caso corregirlos.

Posteriormente, una vez identificadas las diferentes regiones homogéneas, es aconsejable recalcular la medida de la discordancia de cada estación en su región asignada. En aquellos sitios que aquí resulten discordantes habrá que considerar la posibilidad de trasladarlos a otra región, si bien esta opción no prevalecerá sobre las similitudes físicas con el resto de sitios de su región asignada. De hecho, son habituales los eventos que afectan únicamente a algunas estaciones de una determinada región homogénea desde el punto de vista físico, y ello puede dar pie a pensar erróneamente que las estaciones afectadas son discordantes con el resto.

Se concluye insistiendo en la enorme trascendencia de esta primera fase del estudio si se quiere garantizar un análisis riguroso y fiable.

2.3.2.- IDENTIFICACIÓN DE REGIONES HOMOGÉNEAS

De todas las etapas del análisis regional de frecuencias, la identificación de regiones homogéneas es a menudo la más complicada, al requerir de la toma de decisiones subjetivas. El objetivo que se plantea es formar grupos de estaciones que satisfagan aproximadamente la condición de homogeneidad, esto es, que compartan la misma distribución de frecuencias excepto por un factor de escala.

De normal, los datos disponibles a la hora de regionalizar se reducen a las observaciones registradas en las estaciones y a una serie de descriptores que tienen en cuenta sus características físicas (localización geográfica, altitud...). Es importante esta distinción entre observaciones y descriptores, pues Hosking y Wallis (1997) proponen basar la definición de las diferentes regiones en los descriptores físicos, reservando las observaciones para testear la homogeneidad de la regionalización propuesta.

Diversos autores han establecido criterios alternativos para agrupar estaciones. La delimitación según áreas administrativas (Beable y McKerchar, 1982) constituye una de las opciones, si bien a menudo resulta arbitraria y carente de justificación física. En ciertos estudios, sobretudo a pequeña escala, regionalizar por simple inspección subjetiva de las características físicas de las estaciones puede arrojar buenos resultados, tal y como demuestran los trabajos de Schaefer (1990) para las precipitaciones máximas anuales de Washington. El éxito en esta forma de proceder depende de la destreza a la hora de seleccionar y comparar los descriptores físicos más relevantes. Por otro lado, métodos de división basados en criterios objetivos también son posibles; la asignación de una estación en una región u otra en función de si supera o no un umbral establecido para un determinado parámetro físico, o los procesos iterativos hasta conseguir un número de regiones aceptablemente homogéneas, constituyen dos de los casos más habituales.

No obstante, Hosking y Wallis (1997) consideran el análisis cluster como la alternativa más apropiada. Se trata de un método estándar de análisis multivariante por el que se asocia un vector a cada una de las estaciones, que son divididas y agrupadas atendiendo

precisamente a la similitud entre vectores. Los estudios de Acreman y Sinclair (1986) para formar regiones en Escocia, o el trabajo de Guttman (1993) repartiendo 1119 estaciones de Estados Unidos son tan sólo dos ejemplos de los múltiples análisis cluster llevados a cabo hasta la fecha en el ámbito de la hidrología.

En el contexto del análisis regional de frecuencias, hay que tomar en cuenta ciertas consideraciones. Los clusters se forman por agrupación de sitios cuyas características físicas son parecidas, y la mayoría de los algoritmos de clustering miden la similitud de dichas características atendiendo a sus distancias euclídeas en el espacio. Al verse afectadas estas distancias por la escala de medida, es habitual estandarizar cada variable física con el objetivo de conseguir un mismo escenario de dispersión para todas ellas. Sin embargo, es de esperar que ciertas variables condicionen más que otras la forma de la función de distribución, por lo que, pese a la dificultad que ello entraña, se aconseja ponderar y otorgar diferentes pesos a cada una de ellas.

No hay que pasar por alto la idea de que no existe un número de clusters “correcto”, sino que es el área estudiada la que marca la necesidad de agrupar en más o menos regiones. En este sentido, es evidente que clusters con pocas estaciones corren el riesgo de mejorar muy poco la precisión en las estimaciones con respecto a los métodos de análisis local, mientras que regiones de gran tamaño pueden vulnerar el criterio de homogeneidad. Por ello, la elección de algoritmos que tienden a formar clusters de tamaño y variabilidad parecida, como el método de Ward o el enlazado promedio, parece la opción más razonable.

Se destaca además otro aspecto importante. El resultado que arroja un análisis cluster no tiene por qué ser definitivo. Son muchos los ajustes de tipo subjetivo que pueden ser empleados para mejorar la coherencia física del reparto e incluso reducir la heterogeneidad de ciertas regiones. Entre otros:

- mover estaciones de una región a otra;
- suprimir o prescindir de ciertas estaciones;
- subdividir regiones resultantes;
- romper regiones mediante el traslado de estaciones de un cluster a otro;
- combinar regiones entre sí; y
- obtener nuevos datos y redefinir la regionalización.

Establecidas y ajustadas las diferentes regiones, llega el momento de emplear las observaciones registradas en las estaciones para comprobar la hipótesis de homogeneidad. Diferentes tipos de tests han sido propuestos por Dalrymple (1960), Acreman y Sinclair (1986), Wiltshire (1986a,b), Buishand (1989), Chowdhury *et al.* (1991), Lu y Stedinger (1992), Hosking y Wallis (1993), y Fill y Stedinger (1995). La mayoría de ellos implican una cantidad δ que mide un determinado aspecto de la distribución de frecuencias y que es constante para una región homogénea: δ puede ser, por ejemplo, el evento de 10 años escalado por la media (casos de Dalrymple, 1960; Lu y Stedinger, 1992; Fill y Stedinger, 1995), el coeficiente de variación (Wiltshire, 1986a), una combinación de L -CV y L -asimetría (Chowdhury *et al.*, 1991), o el L -CV o la combinación de L -CV, L -asimetría y L -kurtosis (Hosking y Wallis, 1993). Se calculan las estimaciones de δ : $\hat{\delta}^{(i)}$ es la estimación local de la estación i basada en sus observaciones, y $\hat{\delta}^R$ la estimación regional calculada a partir de los datos de todas las

estaciones de la región y asumiendo homogeneidad. Se construye entonces un estadístico S que mide la diferencia entre las estimaciones locales y la estimación regional; una posible elección es

$$S = \sum_{i=1}^N (\hat{\delta}^{(i)} - \hat{\delta}^R)^2. \quad (2.4)$$

El valor observado de S es comparado con la distribución que S tendría si la región fuera verdaderamente homogénea, y este cálculo a menudo implica el asumir una determinada forma para la distribución de frecuencias de la región: desde la Gumbel empleada por Dalrymple (1960) y Fill y Stedinger (1995), por ejemplo, hasta la General de Valores Extremos asumida por Chowdhury *et al.* (1991). Si el valor observado de S queda en la cola de la distribución, se rechaza la hipótesis de homogeneidad por considerarse poco probable que un valor tan extremo de S pueda ser debido al azar.

A pesar de las diferentes alternativas para proceder, en el contexto de un análisis regional de frecuencias, Hosking y Wallis (1997) recomiendan el empleo de tests basados en los L -momentos, asumiendo una distribución Kappa:

- Kappa (KAP). Parámetros (4): ξ (localización), α (escala), k , h .

Su función de distribución se expresa como:

$$F(x) = \left[1 - h \{ 1 - k(x - \xi)/\alpha \}^{1/k} \right]^{1/h} \quad (2.5)$$

Su descripción completa se recoge en el anexo I. Hosking y Wallis (1997) constatan que al tratarse de una función tetraparamétrica resulta menos restrictiva que la función Gumbel o la General de Valores Extremos. Proponen además una medida de la heterogeneidad - validada en una serie de experimentos de simulación tipo Monte Carlo - para evaluar hasta qué punto un grupo de estaciones puede ser tratado como una región homogénea, comparando las variaciones entre los L -momentos muestrales con las que cabría esperar en caso de que lo fuera.

Se puede decir que en una región homogénea todas las estaciones tienen los mismos L -momentos poblacionales. No ocurre lo mismo con los L -momentos muestrales, que son diferentes debido precisamente a la variabilidad que surge del muestreo. Por tanto, la cuestión que se plantea (ver figura 2.6) es averiguar si la dispersión entre los L -momentos muestrales de las estaciones que constituyen la región considerada (datos observados) es mayor que la que se obtendría en caso de homogeneidad (datos simulados).

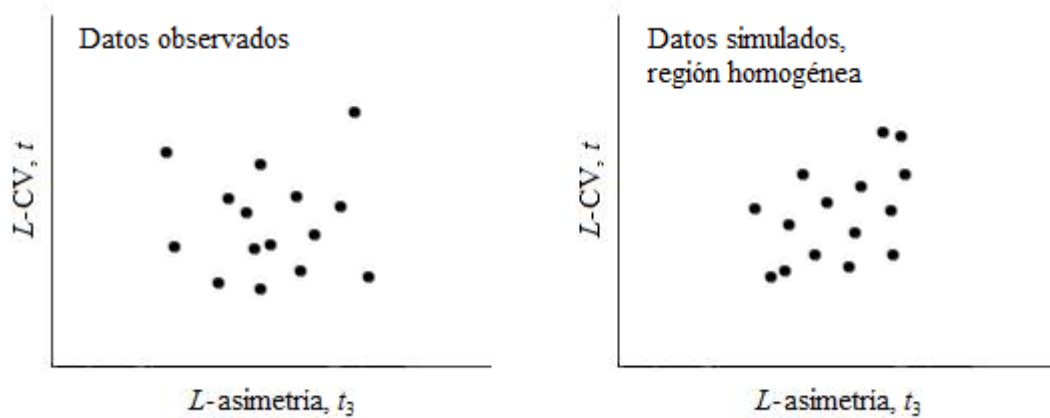


Fig. 2.6. Comparativa entre los datos observados y los datos simulados.

Una evaluación visual de la dispersión entre los L -momentos de las estaciones puede ser obtenida representando gráficamente los valores de L -asimetría frente a los L -CV y frente a los L -kurtosis. El valor numérico de la medida vendría marcado por la distancia media de un punto de una estación al punto que representara la media grupal. Evidentemente, para permitir la mayor variabilidad posible en los L -momentos muestrales, las medias habrían de ser ponderadas proporcionalmente en función de la longitud de la muestra. Como alternativa simple e inmediata para medir la dispersión está la desviación estándar de los L -CV, también ponderada en función de la longitud muestral. El hecho de centrarse en los L -CV se explica porque su variabilidad entre estaciones tiene un efecto mayor que la variabilidad de los L -asimetría y L -kurtosis en la varianza de las estimaciones de la mayoría de los cuantiles.

Por otro lado, para definir lo que cabría esperar en caso de homogeneidad se emplea la simulación repetida de regiones con el mismo número de estaciones y las mismas longitudes muestrales que la real. Gracias a estas simulaciones se obtiene la media y la desviación estándar de la medida de dispersión elegida. Para comparar los valores de dispersión observados con los simulados, un estadístico apropiado es:

$$\frac{(\text{dispersión observada}) - (\text{media simulaciones})}{(\text{desviación estándar simulaciones})} \quad (2.6)$$

Valores positivos y elevados del estadístico indican que los L -momentos observados están más dispersos que los correspondientes a una situación de homogeneidad.

Respecto a la distribución a partir de la cual se generan los datos simulados, Hosking y Wallis (1997) insisten en las ventajas que proporciona la función tetraparamétrica Kappa. Ésta incluye como casos especiales las distribuciones Logística Generalizada, General de Valores Extremos, y General de Pareto, por lo que es capaz de abarcar la mayor parte de las distribuciones empleadas en estudios medioambientales. Sus L -momentos pueden ser escogidos igualándolos a la media regional de L -CV, L -asimetría y L -kurtosis de las observaciones.

Se procede a describir numéricamente la metodología a llevar a cabo. Supóngase que la región propuesta tiene N estaciones, siendo para cada estación i la longitud muestral n_i y los ratios de L -momentos $t^{(i)}$, $t_3^{(i)}$ y $t_4^{(i)}$. Denótese como t^R , t_3^R y t_4^R los ratios medios regionales de L -CV, L -asimetría y L -kurtosis, ponderados proporcionalmente en función de la longitud muestral; por ejemplo

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i}. \quad (2.7)$$

Se calcula la desviación estándar ponderada de los L -CV locales,

$$V = \left\{ \sum_{i=1}^N n_i (t^{(i)} - t^R)^2 / \sum_{i=1}^N n_i \right\}^{1/2}. \quad (2.8)$$

Se ajusta una distribución Kappa a los ratios medios regionales 1 , t^R , t_3^R y t_4^R . El escoger el valor 1 como media regional se explica por el hecho de que la media de los valores reescalados q_{ij} de cada estación es 1, y por tanto la media regional de esas medias también es 1. Se realiza un número M de simulaciones de una región con N sitios, como la original, cada una de ellas teniendo la función Kappa como distribución de frecuencias. Para cada simulación se calcula V . Se determina además la media, μ_V , y la desviación estándar, σ_V , de los M valores de V . La medida de heterogeneidad se calcula como

$$H = \frac{(V - \mu_V)}{\sigma_V}. \quad (2.9)$$

La región se declara heterogénea si el valor de H es suficientemente elevado. Hosking y Wallis (1997) sugieren considerarla como “aceptablemente homogénea” si $H < 1$, “posiblemente heterogénea” si $1 \leq H < 2$, y “definitivamente heterogénea” si $H \geq 2$. Valores negativos indican una dispersión de los L -CV muestrales menor que la esperada en caso de región homogénea con distribuciones locales independientes, siendo la existencia de correlación positiva entre estaciones su causa más probable.

Es necesario remarcar ciertas consideraciones. En muchos estudios, la comparación entre valores observados y simulados se ha llevado a cabo de manera informal, empleando únicamente una o dos simulaciones (Hosking, 1990; Pearson, 1991, 1993; Pilon, Adamowski y Alila, 1991; Wallis, 1993). Sin embargo, el uso del estadístico H con un número mayor de repeticiones constituye una alternativa más fiable y menos subjetiva. Se considera adecuado un valor M de 500 simulaciones, requiriéndose de valores mayores en caso de que H resulte cercano a los límites 1 y 2.

Por otro lado, se pueden presentar situaciones en las que no sea posible ajustar la función Kappa a los ratios regionales ponderados. Esto ocurre cuando t_4^R es muy elevado con respecto a t_3^R . Para estos casos, Hosking y Wallis (1997) proponen el empleo en la simulación de la distribución Logística Generalizada, un caso particular de Kappa con parámetro $h = -1$.

La posibilidad de construir el estadístico de heterogeneidad H recurriendo a otras medidas de dispersión diferentes a la planteada en la ecuación 2.8 también puede ser considerada. No obstante, se insiste en el mayor poder discriminatorio de los L -CV frente a los L -asimetría y L -kurtosis, y en su mayor capacidad para comparar estaciones mediante el procedimiento del índice de avenida.

En lo que respecta al uso de la medida de heterogeneidad H , ésta ha de ser empleada para evaluar precisamente el grado de heterogeneidad de una región propuesta. Siempre que dicha región no resulte “aceptablemente homogénea”, hay que valorar la opción de redefinir la regionalización mediante un nuevo análisis cluster o por medio de los ajustes subjetivos descritos. Sin embargo, al igual que ocurre con la medida de la discordancia, son frecuentes los casos en los que existe una aparente heterogeneidad causada por la presencia de estaciones que, pese a poseer similitudes físicas evidentes con su región, reúnen valores atípicos. En estas situaciones, el argumento físico prevalece sobre el estadístico puesto que, durante el periodo que abarcan las muestras, determinados eventos han podido ocurrir únicamente en ciertas estaciones de la región. Y es en estos casos donde el análisis regional de frecuencias aporta su mayor potencial.

Pese a que el estadístico H se construye como un test de significación de la hipótesis de homogeneidad, Hosking y Wallis (1997) recomiendan no emplearlo en ese sentido. No hay que olvidar el hecho de que los cinco supuestos que asume el método del índice de avenida nunca se cumplen estrictamente, sino que a menudo es de esperar la existencia de cierta correlación interestacional en las observaciones. No obstante, a efectos prácticos, incluso con una heterogeneidad moderada se pueden conseguir estimaciones precisas.

Se concluye enfatizando nuevamente en la importancia de basar la definición de las diferentes regiones en los descriptores físicos, reservando las observaciones para testear la homogeneidad de la regionalización propuesta.

2.3.3.- ELECCIÓN DE LA DISTRIBUCIÓN DE FRECUENCIA

Definidas las regiones, se procede a determinar la función de distribución más apropiada para cada una de ellas. En general, se puede decir que la homogeneidad total es imposible. Toda región es siempre ligeramente heterogénea, y no existe una distribución “verdadera” que la pueda caracterizar perfectamente. Por tanto, el objetivo se centra en encontrar aquella que proporciona las estimaciones más precisas para cada estación.

Es importante remarcar que la distribución elegida no tiene por qué ser la que más se aproxime a las observaciones. El hecho de existir una función con un buen ajuste a los datos no garantiza que, en ella, los valores futuros de la variable vayan a ser coherentes con los del pasado, pues a menudo derivan de procesos físicos propensos a originar *outliers* alejados del resto. En este sentido, es preferible anteponer la elección de un modelo robusto en estimaciones a la selección de la distribución que mejor ajusta.

Por lo general, existe un rango de periodos de retorno de interés para los cuales se requiere estimar cuantiles. En el estudio de eventos extremos, como precipitaciones,

avenidas o sequías, son de particular relevancia las estimaciones que recaen en la cola de la distribución de frecuencia. Ésta y otras muchas consideraciones han de ser tenidas en cuenta a la hora de seleccionar.

Para el análisis regional de frecuencias, Hosking y Wallis (1997) recomiendan la elección entre las siguientes cinco distribuciones:

- Logística Generalizada (GLO). Parámetros (3): ξ (localización), α (escala), k (forma).

Su función de distribución se expresa como:

$$F(x) = \frac{1}{(1 + e^{-y})} \quad \text{siendo } y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases} \quad (2.10)$$

- General de Valores Extremos (GEV). Parámetros (3): ξ (localización), α (escala), k (forma).

Su función de distribución se expresa como:

$$F(x) = e^{-e^{-y}} \quad \text{siendo } y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases} \quad (2.11)$$

- General de Pareto (GPA). Parámetros (3): ξ (localización), α (escala), k (forma).

Su función de distribución se expresa como:

$$F(x) = 1 - e^{-y} \quad \text{siendo } y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases} \quad (2.12)$$

- LogNormal (LN3). Parámetros (3): ξ (localización), α (escala), k (forma).

Su función de densidad se expresa como:

$$f(x) = \frac{e^{ky - y^2/2}}{\alpha\sqrt{2\pi}} \quad \text{siendo } y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases} \quad (2.13)$$

- Pearson tipo III (PE3). Parámetros (3): μ (localización), σ (escala), γ (forma).

Siendo $\alpha = 4/\gamma^2$, $\beta = \frac{1}{2}\sigma|\gamma|$, $\xi = \mu - 2\sigma/\gamma$, su función de distribución se expresa como:

$$F(x) = \begin{cases} G\left(\alpha, \frac{x-\xi}{\beta}\right) / \Gamma(\alpha), & \gamma > 0 \\ 1 - G\left(\alpha, \frac{\xi-x}{\beta}\right) / \Gamma(\alpha), & \gamma < 0 \end{cases} \quad (2.14)$$

siendo $\Gamma(\cdot)$ la función Gamma y $G(\cdot)$ la función Gamma incompleta.

Sus descripciones con mayor detalle se recogen en el anexo I. Hosking y Wallis (1997) justifican la elección entre estas cinco distribuciones por el hecho de poseer tres parámetros - lo cual, en un análisis regional, facilita el ajuste al disponerse de observaciones de varias estaciones - y de incluir entre sus casos particulares las distribuciones más empleadas en el estudio de variables ambientales.

Pese a ello, son muchas las familias de distribuciones candidatas a ser empleadas en el ajuste de las observaciones. Su idoneidad de cara al análisis puede ser evaluada atendiendo a su capacidad para reproducir las características de los datos que se requieren en el modelo. Para cualquier aplicación, es recomendable considerar ciertas propiedades que facilitan la selección de la más adecuada:

- Límite superior de la distribución. Algunas magnitudes físicas poseen un límite superior. Dicho límite no adopta un valor concreto y conocido, sino que existen cotas numéricas de la variable que se asumen como físicamente imposibles de alcanzar. Valga como ejemplo una velocidad de viento de 600 km/h, o una precipitación instantánea de 500 mm/h. Es evidente que ambos valores jamás pueden ocurrir. Ante estos casos, algunos autores (Boughton, 1980; Laursen, 1983) han establecido la necesidad de recurrir a distribuciones con límite superior. Sin embargo, Hosking y Wallis (1997) consideran este argumento equivocado, pues la imposición del límite puede comprometer la precisión en las estimaciones de los cuantiles correspondientes a los periodos de retorno de interés (sobre todo los más elevados). No hay duda, además, de que la presencia de un límite hace más complicado el ajuste a las observaciones en la cola de la distribución. Por tanto, su imposición ha de plantearse siempre y cuando la precisión de las estimaciones y el ajuste no se vean amenazados.
- Cola superior de la distribución. En la mayoría de estudios, la estimación de la cola superior de la distribución es de gran interés, pues su peso - esto es, el comportamiento de la función de densidad $f(x)$ cuando x aumenta - determina la manera en que los cuantiles crecen a medida que el periodo de retorno es extrapolado más allá del rango de datos. La tabla 2.3 refleja los pesos de la cola de las distribuciones más habituales. Cuando no hay razón para asumir que un solo tipo de cola es el apropiado, es recomendable el empleo de varias funciones candidatas que cubran el rango de los diferentes pesos.

Tabla 2.3. Pesos de las colas superiores de las distribuciones más comunes.

Forma de $f(x)$ para x elevados	Distribuciones
x^{-A}	General de Valores Extremos, General de Pareto y Logística Generalizada con parámetro $k < 0$.
$x^{-A \log x}$	LogNormal con asimetría positiva.
$\exp(-x^A)$, $0 < A < 1$	Weibull con parámetro $\lambda < 1$.
$x^A e^{-Bx}$	Pearson tipo III con asimetría positiva.
$\exp(-x)$	Exponencial, Gumbel.
$\exp(-x^A)$, $A > 1$	Weibull con parámetro $\lambda > 1$.
Límite superior finito	General de Valores Extremos, General de Pareto y Logística Generalizada con parámetro $k > 0$; LogNormal y Pearson tipo III con asimetría negativa.

Nota: los pesos de las colas se ordenan de más pesado a más ligero.
 A y B denotan constantes positivas arbitrarias.

- Forma del cuerpo de la distribución. La mayor parte de las distribuciones empleadas en análisis de tipo estadístico se caracterizan por poseer una función de densidad con un pico más o menos marcado. No obstante, para ciertas variables ambientales se pueden plantear otras formas de distribución. Por ejemplo, los estudios de caudales máximos anuales a menudo se cimientan en valores procedentes de crecidas originadas por deshielo o eventos de tormenta sucesivos. En estas situaciones en las que las observaciones se derivan de diferentes causas, una distribución combinada puede ser asumida bajo la forma:

$$F(x) = pG_1(x) + (1-p)G_2(x), \quad (2.15)$$

donde $G_1(x)$ y $G_2(x)$ son las funciones acumuladas de distribución de los datos procedentes de dos causas distintas, y p la proporción de observaciones derivados de la primera causa. Pese a que entraña mayor dificultad, Hosking y Wallis (1997) plantean la posibilidad de recurrir a ella en aquellos análisis de tipo regional basados en L -momentos que así lo requieran.

- Cola inferior de la distribución. Se aplican las mismas consideraciones que para la cola superior; en muchos casos, es aconsejable considerar un rango de distribuciones con diferentes pesos de cola. No obstante, si el interés se centra en la cola superior, la forma de la cola inferior resulta irrelevante. A modo de ejemplo, en estudios de caudales máximos anuales, una distribución exponencial

puede aproximar correctamente los cuantiles de la cola superior (Damazio y Kelman, 1986) aun cuando la cola inferior guarda pocas similitudes con la forma de las observaciones.

- Límite inferior de la distribución. También se aplican las mismas consideraciones que para el límite superior; incluso cuando éste existe, rara vez resulta eficiente tratar de estimarlo explícitamente. Sin embargo, contrariamente al límite superior, el inferior a menudo es conocido e igual a cero. De ahí que en aquellas situaciones en que los cuantiles de interés son cercanos a cero pueda merecer la pena su imposición mediante el uso de funciones tipo Wakeby:

Wakeby (WAK). Parámetros (5): ξ (localización), α , β , γ , δ .

Su función cuantil se expresa como:

$$x(F) = \xi + \frac{\alpha}{\beta} \{1 - (1 - F)^\beta\} - \frac{\gamma}{\delta} \{1 - (1 - F)^{-\delta}\} \quad (2.16)$$

Su descripción completa se recoge en el anexo I. Las funciones General de Pareto o Pearson tipo III constituyen otras opciones para estos casos. En el resto de aplicaciones, caso de las precipitaciones anuales, rara vez resulta útil; a menudo se obtienen mejores resultados ajustando a una distribución en la que se ha fijado un límite superior a cero o se ha prescindido de él.

- Valores cero. Cuando la variable de estudio contiene valores cero y además la estimación de cuantiles en la cola inferior de la distribución es importante, el empleo de funciones transformadas que permitan una proporción de valores nulos puede resultar útil. No obstante, Hosking y Wallis (1997) destacan la dificultad de basar este tipo de ajustes en los L -momentos. Como alternativa, proponen el empleo de una distribución mixta bajo la forma:

$$F(x) = \begin{cases} 0, & x < 0, \\ p + (1 - p)G(x), & x \geq 0. \end{cases} \quad (2.17)$$

Aquí p es la probabilidad de un valor cero y $G(x)$ la función acumulada de distribución de los valores nulos. El parámetro p puede ser estimado por la proporción de valores cero en los datos de la región, y la distribución $G(x)$ se ajusta mediante los L -momentos regionales de los valores nulos. Diversos estudios han empleado este procedimiento con éxito (Guttman, Hosking y Wallis, 1993).

Por otro lado, es evidente que en ocasiones se presentan razones teóricas suficientes como para optar por una familia particular de distribuciones para un tipo de datos concreto. Desde los estudios de Gumbel (1958) se ha asumido que en el análisis de eventos extremos (máximos anuales de precipitación o caudal, por ejemplo) es imprescindible el empleo de distribuciones de valores extremos. No obstante, Hosking y Wallis (1997) remarcan la necesidad de contrastar siempre su idoneidad frente al resto de candidatas, sin descartar de antemano ninguna posibilidad.

El número de parámetros de cada distribución también es de suma relevancia. Se puede afirmar que las funciones biparamétricas son capaces de aportar estimaciones precisas siempre y cuando la verdadera distribución de los datos sea muy cercana a la de ajuste. En caso de que su forma difiera, especialmente en las colas, los cuantiles estimados más extremos se ven comprometidos al aumentar su error. De ahí que el empleo de funciones con mayor número de parámetros pueda resultar crucial en muchas de las situaciones que se presentan. Una de las ventajas del análisis regional de frecuencias es que la estimación de distribuciones de tres o más parámetros es, a priori, mucho más fiable que la que se puede esperar del método clásico local. A modo de ejemplo, el uso de la función pentaparamétrica Wakeby cuando en cada estación se dispone de 20 observaciones se considera una posibilidad razonable. Por ello, Hosking y Wallis (1997) establecen las distribuciones de tres a cinco parámetros como las más apropiadas.

Se insiste en el hecho de que la finalidad última del análisis regional de frecuencias no es conseguir el mejor ajuste, sino estimar cuantiles procedentes de aquella distribución de la que surgirán valores futuros de la variable. Si son varias las funciones que ajustan correctamente, se opta por seleccionar la más robusta, esto es, la más capacitada para proporcionar buenas estimaciones aun cuando los valores futuros proceden de una distribución diferente a la ajustada.

Son muchas las alternativas disponibles para evaluar la bondad de ajuste de una determinada distribución a una muestra de observaciones: desde el empleo de diagramas cuantil-cuantil, por ejemplo, hasta el uso de los tests de chi-cuadrado o Kolmogorov-Smirnov, bien empleados en buena parte de los estudios estadísticos. Sin embargo, en el contexto del análisis regional de frecuencias, Hosking y Wallis (1997) proponen trabajar directamente sobre los L -momentos regionales. Definen además una medida de bondad de ajuste - validada en una serie de experimentos de simulación tipo Monte Carlo - para evaluar hasta qué punto una distribución concreta se aproxima correctamente a los datos, y seleccionar de entre las candidatas aquella que proporciona el mejor ajuste.

Asúmase una región aceptablemente homogénea. Se puede decir que los ratios de L -momentos de las estaciones que la forman quedan resumidos en los valores medios regionales. Los parámetros de localización y escala de la distribución evaluada se eligen igualándolos a la media regional y al L -CV regional. Así, la bondad de ajuste juzga cuánto se aproximan los L -asimetría y L -kurtosis de la distribución a los valores regionales de L -asimetría y L -kurtosis. Pueden emplearse igualmente L -momentos de orden superior, si bien en principio no es necesario.

Para obtener la medida de bondad de ajuste, se procede de la siguiente manera. Se asume por conveniencia que la distribución candidata es la General de Valores Extremos (GEV), de tres parámetros, siendo los L -asimetría y L -kurtosis muestrales insesgados. La distribución GEV, ajustada por el método de los L -momentos, tiene un valor de L -asimetría igual al valor medio regional de dicho coeficiente. Se juzga entonces la calidad del ajuste por la diferencia entre el L -kurtosis de la distribución GEV, τ_4^{GEV} , y el valor medio regional de L -kurtosis, t_4^R (ver figura 2.7).

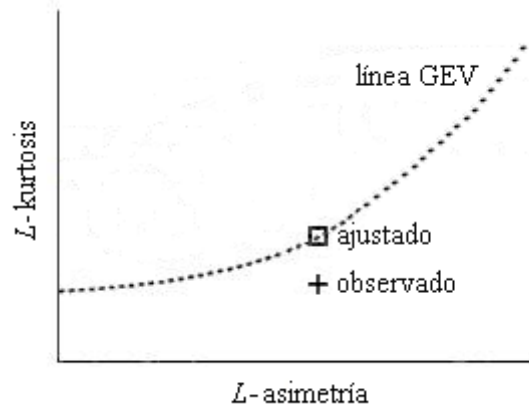


Fig. 2.7. Representación gráfica del concepto de bondad de ajuste.

Para evaluar si la diferencia es significativa, se compara ésta con la variabilidad de t_4^R . Denotándose como σ_4 la desviación estándar de t_4^R , se calcula mediante repetidas simulaciones de una región homogénea cuyas estaciones siguen una distribución GEV con longitudes muestrales similares a las observadas. Entonces se establece:

$$Z^{GEV} = (t_4^R - \tau_4^{GEV}) / \sigma_4 \quad (2.18)$$

como medida de bondad de ajuste; valores pequeños del estadístico Z^{GEV} indican que la distribución GEV es la adecuada para la región.

Una dificultad que presenta el procedimiento descrito es que cada distribución candidata implica la realización de un conjunto de simulaciones para obtener σ_4 . Sin embargo, a efectos prácticos, es razonable asumir el mismo valor para todas ellas, dado que su L -asimetría es igual y ello garantiza valores de σ_4 más o menos similares. Bajo esta suposición, Hosking y Wallis (1997) recomiendan el empleo de la función tetraparamétrica Kappa aprovechando las mismas simulaciones llevadas a cabo en el cálculo de la medida de heterogeneidad.

Anteriormente se han asumido los L -asimetría y L -kurtosis como insesgados. Esta aproximación puede ser buena para t_3 , pero no para t_4 cuando las longitudes muestrales son cortas ($n_i \leq 20$) o el valor poblacional de L -asimetría es elevado ($\tau_3 \geq 0.4$). El problema se soluciona mediante la corrección del sesgo de t_4 ; así, se compara el coeficiente L -kurtosis ajustado, τ_4^{GEV} , no con el valor regional t_4^R , sino con su versión corregida $t_4^R - B_4$, donde B_4 expresa el sesgo de la media regional de L -kurtosis en regiones con el mismo número de estaciones y longitudes muestrales que la observada, y se calcula con las mismas simulaciones empleadas para obtener σ_4 .

Se describe a continuación el procedimiento numérico a seguir. Supóngase una región de N estaciones, cada estación i con longitud muestral n_i y ratios de L -momentos $t^{(i)}$, $t_3^{(i)}$ y $t_4^{(i)}$. Denótense como t^R , t_3^R y t_4^R los ratios medios regionales de L -CV, L -asimetría y

L -kurtosis, ponderados proporcionalmente en función de la longitud muestral (como en la ecuación 2.7)

Considérese el conjunto de distribuciones candidatas de tres parámetros descrito anteriormente: Logística Generalizada (GLO), General de Valores Extremos (GEV), General de Pareto (GPA), LogNormal (LN3) y Pearson tipo III (PE3).

Se ajusta cada distribución por separado a los ratios medios regionales $1, t^R$ y t_3^R . Se denota como τ_4^{DIST} el coeficiente L -kurtosis de la distribución ajustada, donde DIST hace referencia a GLO, GEV, GPA, etc.

Paralelamente, se ajusta una distribución Kappa a los ratios medios regionales $1, t^R, t_3^R$ y t_4^R , y se realiza un número M de simulaciones de una región con N sitios, como la original, cada una de ellas teniendo la función Kappa como distribución de frecuencias. Se insiste en la posibilidad de emplear las mismas simulaciones llevadas a cabo para el cálculo de la heterogeneidad. Para la m -ésima región simulada, se determina el valor regional de L -asimetría, $t_3^{[m]}$, y de L -kurtosis, $t_4^{[m]}$.

Se calcula el sesgo de t_4^R ,

$$B_4 = M^{-1} \sum_{m=1}^M (t_4^{[m]} - t_4^R), \quad (2.19)$$

la desviación estándar de t_4^R ,

$$\sigma_4 = \left[(M-1)^{-1} \left\{ \sum_{m=1}^M (t_4^{[m]} - t_4^R)^2 - M \cdot B_4^2 \right\} \right]^{1/2}, \quad (2.20)$$

y, para cada distribución candidata, la medida de bondad de ajuste

$$Z^{DIST} = (\tau_4^{DIST} - t_4^R + B_4) / \sigma_4. \quad (2.21)$$

El ajuste se declara adecuado si Z^{DIST} es suficientemente cercano a cero; un criterio adecuado, según Hosking y Wallis (1997), es establecer el límite $|Z^{DIST}| \leq 1.64$.

No obstante, hay que tomar en cuenta varias consideraciones, pues dicho criterio es arbitrario en algunos casos. El estadístico Z adquiere forma de test de significación de bondad de ajuste y sigue una distribución Normal estandarizada sólo bajo ciertas suposiciones. Así, $|Z| \leq 1.64$ correspondería a la aceptación de la distribución hipotética a un nivel de confianza del 90%. Sin embargo, esas suposiciones incluyen dos requisitos difícilmente alcanzables en la práctica: que la región sea exactamente homogénea y que no exista dependencia entre estaciones. Por eso Hosking y Wallis (1997) recomiendan el uso de Z como una aproximación al ajuste y no como un test formal.

Por otro lado, se puede afirmar que el criterio $|Z| \leq 1.64$ es poco fiable en situaciones particulares en las que existe fuerte correlación serial o interestacional. La existencia de correlación tiende a aumentar la variabilidad de t_4^R , de forma que al no incluirse ésta en las simulaciones de la región mediante la función Kappa, la estimación de σ_4 resulta reducida y los valores de Z demasiado elevados. Ello deriva en la indicación falsa de que el ajuste no es bueno.

En lo que respecta al uso de la medida de bondad de ajuste, se establecen como aceptables todas las distribuciones candidatas que satisfacen la condición $|Z| \leq 1.64$. Para todas ellas se calcula por separado la curva regional de frecuencia. Si éstas son aproximadamente iguales, cualquiera de las distribuciones es adecuada, si bien interesa optar por el modelo más robusto frente a posibles errores cometidos en la definición de la región. En caso de que las curvas regionales difieran, puede existir un problema de escasez de datos que conduce a diferencias estadísticamente insignificantes pero operacionalmente importantes entre las distribuciones candidatas. Para esas situaciones en las que no es posible identificar de manera fiable el mejor modelo, la robustez también adquiere especial trascendencia, y el empleo de funciones tipo Kappa (cuatro parámetros) o Wakeby (cinco parámetros) se plantea como alternativa aconsejable.

Puede ocurrir que ninguna de las distribuciones cumpla el criterio $|Z| \leq 1.64$, siendo la presencia de un número elevado de estaciones por región o de series de observaciones demasiado largas la causa más probable. En estas circunstancias, σ_4 es reducida y el estadístico Z puede ser elevado aun cuando los valores regionales de L -asimetría y L -kurtosis se aproximan a alguna de las distribuciones candidatas. Si el punto (t_3^R, t_4^R) recae entre varias distribuciones cuyas curvas regionales de frecuencia son aproximadamente iguales, se presenta el problema contrario al anterior: superabundancia de datos que conduce a diferencias estadísticamente significativas pero operacionalmente poco importantes. Aquí, reclasificar el ajuste de cualquiera de las candidatas se convierte en primordial. Cuando el punto no se ubica entre las distribuciones, el empleo de funciones Kappa o Wakeby es más recomendable.

Por último, en regiones que no son aceptablemente homogéneas, no existe razón para dar por supuesto que una función de distribución pueda caracterizarlas adecuadamente, pero tampoco para descartar su existencia. Una heterogeneidad moderada también es capaz de arrojar estimaciones precisas de cuantiles por el método regional; aquí, el establecer un modelo robusto a la extrapolación local se considera crucial, y nuevamente las distribuciones Kappa y Wakeby resultan útiles por tener mayor número de parámetros.

Para regiones claramente heterogéneas, Hosking y Wallis (1997) recomiendan anteponer los L -momentos locales a los regionales de cara a discriminar entre las distribuciones candidatas. Es evidente que, ante la existencia de heterogeneidad, basar el ajuste en los valores medios regionales carece de sentido, pues no son representativos de cada estación de la región. Aquí, el empleo de la función pentaparamétrica Wakeby es igualmente recomendable.

2.3.4.- ESTIMACIÓN DE LOS CUANTILES DE FRECUENCIA

Establecidas las diferentes regiones y escogida la distribución de frecuencias más apropiada para cada una de ellas, se procede a determinar sus parámetros de cara a la estimación de los cuantiles de interés. Se recuerda que la relación entre las distribuciones de las estaciones que componen una región aceptablemente homogénea es la justificación del análisis regional de frecuencias. Es precisamente ésta la que puede permitir que los cuantiles estimados por combinación de los datos de todos los sitios sean más precisos que los obtenidos analizando cada estación por separado.

Son muchos los métodos existentes para determinar los parámetros de la distribución ajustada. Para describirlos, se recupera la notación planteada en la introducción al método del índice de avenida (ecuación 1.1). Supónganse que la región tiene N estaciones, cada estación i con un tamaño muestral n_i y datos observados Q_{ij} , $j = 1, \dots, n_i$. Sea $Q_i(F)$, $0 < F < 1$ la función cuantil de la distribución de frecuencia en la estación i . Para una región homogénea, se define:

$$Q_i(F) = \mu_i q(F), \quad i = 1, \dots, N, \quad (2.22)$$

donde μ_i es el factor de escala local de la estación i , esto es, el índice de avenida de dicha estación, siendo $\hat{\mu}_i$ su estimador; y $q(F)$ es la curva regional de frecuencia. Los datos reescalados adimensionalizados son $q_{ij} = Q_{ij} / \hat{\mu}_i$, $j = 1, \dots, n_i$, $i = 1, \dots, N$.

El procedimiento año-estación combina los datos reescalados q_{ij} de todas las estaciones en una sola muestra a la que se ajusta la distribución apropiada como si ésta fuera aleatoria. A día de hoy raramente se usa, dado que en muchos casos resulta inapropiado asumir esa combinación de datos como muestra aleatoria; es obvio que cuando los estimadores $\hat{\mu}_i$ presentan diferente precisión por variar las longitudes muestrales de las estaciones, los datos reescalados de cada sitio no están idénticamente distribuidos.

Por otro lado, la aproximación por el método de máxima verosimilitud trata la ecuación 2.22 como un modelo estadístico perfectamente especificado por los N factores de escala μ_i , $i = 1, \dots, N$, y los p parámetros desconocidos de la curva regional de frecuencia $q(F)$. Los estudios de Boes, Heo y Salas (1989) y Buishand (1989) corroboran la posibilidad de estimar estos $N + p$ parámetros por máxima verosimilitud.

No obstante, el método del índice de avenida descrito en la introducción emplea los descriptores estadísticos de los datos de cada estación para combinarlos entre sí de forma ponderada y obtener las estimaciones regionales definidas en la ecuación 1.2. Cuando los estadísticos utilizados son los ratios de L -momentos, el procedimiento recibe el nombre de “algoritmo regional de L -momentos”. Frente al método de máxima verosimilitud, se presenta como una alternativa más razonable de cara a combinar la información de los diferentes sitios que componen la región. Además, su cálculo tiende a ser más sencillo por no requerir del tratamiento simultáneo de todos los datos disponibles: los ratios medios regionales de L -momentos, computados en el propio algoritmo, reúnen y sintetizan las características de la colección de datos de la región.

Por todo ello, Hosking y Wallis (1997) consideran el método del índice de avenida implementado en el algoritmo regional de L -momentos como la opción más conveniente y eficiente para estimar los parámetros de la distribución y los cuantiles de interés. Sin olvidar que asume la independencia de las observaciones tanto dentro de las series como entre estaciones, siempre que exista dependencia y no sea elevada la precisión de las estimaciones no tiene por qué verse afectada.

El objetivo que se plantea es ajustar a los datos de las estaciones de la región homogénea una distribución regional de frecuencias que se particulariza en cada una de ellas mediante un factor de escala (el índice de avenida). Después, la estimación de cuantiles se lleva a cabo por medio de la distribución local.

La distribución se ajusta por el método de los L -momentos; sus parámetros se estiman igualando los L -momentos poblacionales de la distribución a los L -momentos muestrales calculados a partir de las observaciones. Asumida la región como homogénea, los ratios muestrales de L -momentos obtenidos a partir de los datos reescalados q_{ij} de las diferentes estaciones se combinan para obtener los ratios medios regionales de L -momentos. Con el fin de permitir la mayor variabilidad posible en los ratios de muestras pequeñas, los valores medios se ponderan proporcionalmente a la longitud muestral.

Para mayor simplicidad se asume que el índice de avenida es la media de la distribución de frecuencia en cada sitio, estimada por el valor medio de sus observaciones. Entonces la media de los valores reescalados q_{ij} de cada estación es 1, y por tanto la media regional de las medias también es 1. Por otro lado, para cada sitio los ratios de L -momentos t y t_r , $r \geq 3$, son los mismos independientemente de si se calculan a partir de los datos reescalados q_{ij} o de los originales Q_{ij} , por lo que computar los primeros no es necesario.

Se procede a describir numéricamente la metodología a llevar a cabo. Sea la región de N estaciones, cada estación i con longitud muestral n_i , media muestral $l_1^{(i)}$, y ratios de L -momentos $t^{(i)}, t_3^{(i)}, t_4^{(i)}, \dots$. Denótese como t^R, t_3^R, t_4^R, \dots , los ratios medios regionales de L -momentos, ponderados proporcionalmente en función de la longitud muestral:

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i}, \quad (2.23)$$

$$t_r^R = \frac{\sum_{i=1}^N n_i t_r^{(i)}}{\sum_{i=1}^N n_i}, \quad (2.24)$$

Se establece el valor de la media regional en 1, esto es, $l_1^R = 1$.

Se ajusta la distribución igualando sus ratios de L -momentos $\lambda_1, \tau, \tau_3, \tau_4, \dots$, a los ratios medios regionales de L -momentos $l_1^R, t^R, t_3^R, t_4^R, \dots$, calculados. Se denota como $\hat{q}(\cdot)$ la función cuantil de la distribución regional de frecuencia ajustada.

Las estimaciones de cuantiles en la estación i se obtienen por combinación de las estimaciones de μ_i y $q(\cdot)$. Así, el cuantil de probabilidad de no excedencia F es:

$$\hat{Q}_i(F) = l_1^{(i)} \hat{q}(F). \quad (2.25)$$

Pese a que Hosking y Wallis (1997) recomiendan el empleo del método propuesto, de él se derivan multitud de variantes. Por ejemplo, es posible asumir como índice de avenida un valor diferente a la media local de las observaciones. La mediana o un cuantil de frecuencia concreto constituyen otras opciones. Del mismo modo, el empleo de L -momentos en lugar de ratios de L -momentos (Stedinger *et al.*, 1992) es perfectamente viable, si bien la estimación de los cuantiles puede quedar sujeta a mayores imprecisiones.

Por otra parte, en todo análisis estadístico, los valores estimados se enmarcan dentro de un rango de incertidumbre que es necesario determinar si se quiere maximizar su utilidad. Se puede afirmar que evaluar la precisión de las estimaciones resulta fundamental. En métodos tradicionales esto se consigue definiendo intervalos de confianza, asumiendo a menudo la satisfacción de todas las suposiciones de las que parte el modelo. Para un análisis regional de frecuencias basado en el algoritmo regional de L -momentos, también es posible la construcción de los intervalos de confianza de las estimaciones en la región homogénea, al menos como una aproximación cuando los L -momentos muestrales pueden considerarse normalmente distribuidos. Sin embargo, la precisión evaluada de este modo tiende a reflejar únicamente el rigor de los supuestos asumidos por el modelo, tal y como Rosbjerg y Madsen (1995) demostraron; en líneas generales, es evidente que cuando se calcula la precisión de un método partiendo de la base de que sus suposiciones se cumplen, aquél que asume los supuestos más estrictos es propenso a arrojar los resultados más precisos.

Por ello, en el contexto del análisis regional, los intervalos de confianza son de poca utilidad, pues rara vez se puede asegurar que el modelo empleado es del todo correcto, esto es, que se cumplen todos sus supuestos. No obstante, según Hosking y Wallis (1997) el algoritmo regional de L -momentos resulta útil aun cuando las suposiciones no se satisfacen totalmente. Se requiere entonces una forma de evaluar la precisión que tenga en cuenta la esperable dependencia interestacional de las observaciones, y la posibilidad de que la región sea moderadamente heterogénea o la distribución de frecuencias no esté perfectamente caracterizada.

En este sentido, se establece la simulación tipo Monte Carlo como la mejor opción. Las simulaciones han de ser ejecutadas atendiendo a las características de los datos con los que se calculan las estimaciones, por lo que la región simulada debe tener igual número de estaciones, longitudes muestrales y ratios medios regionales de L -momentos que la real. Es adecuado, además, incluir heterogeneidad, errores de especificación y dependencia entre observaciones, o una combinación de ellas en caso de que existan.

Los ratios de L -momentos locales deben ser elegidos de forma que la región simulada alcance una heterogeneidad consistente con las medidas de heterogeneidad reales calculadas a partir de los datos. De ahí que diversas simulaciones preeliminares puedan ser llevadas a cabo hasta conseguir la variabilidad necesaria en dichos ratios, requiriéndose de cierta arbitrariedad y decisión subjetiva al ser varios los patrones posibles.

Un aspecto importante es que la variación interestacional de los ratios de L -momentos de la región simulada ha de ser siempre menor que la de la observada, pues la variabilidad del muestreo implica una mayor dispersión en los L -momentos muestrales que en los poblacionales. Particularmente incorrecto resulta el empleo de los ratios observados como si de los poblacionales de la región simulada se tratara; ello conduce a valores de heterogeneidad más altos que los de la región real.

Las distribuciones de frecuencia individuales de cada estación deben ser escogidas según las medidas de bondad de ajuste obtenidas a partir de los datos. Si son varias las que ajustan correctamente, cualquiera de ellas es apropiada para ser empleada en la simulación; en su defecto, distribuciones más flexibles como Kappa o Wakeby pueden ser utilizadas.

Si la dependencia entre estaciones se aprecia como un problema, puede ser incluida en las simulaciones modificando el algoritmo. Una manera conveniente de estructurarla se plantea asumiendo que si cada distribución de frecuencia local se transformara en una distribución Normal – llámese esta transformación T – entonces la distribución conjunta de los N sitios sería Normal multivariada. Hosking y Wallis (1987) procedieron así en el estudio de caudales máximos anuales en Gran Bretaña, obteniendo buenos resultados. La generación de datos implica aquí los siguientes pasos: generar una matriz R de correlación entre estaciones; generar un vector aleatorio x de distribución Normal multivariada y covarianza la matriz R ; y aplicar la inversa de la transformación T para obtener los datos de la distribución marginal requerida.

Evidentemente, la matriz de correlación R empleada en la simulación de datos ha de ser coherente con las observaciones. Cuando no existe un patrón concreto de correlación entre las estaciones de la región, es razonable considerarlas equicorrelacionadas; así, el valor de correlación entre los sitios i y j es $\rho_{ij} = \rho$ para $i \neq j$, y la matriz R tiene la forma:

$$R = \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix}. \quad (2.26)$$

El valor ρ se estima por la media de las correlaciones entre cada par de estaciones. Sea Q_{ik} el dato del sitio i en el punto temporal k . La correlación entre los sitios i y j viene determinada por

$$r_{ij} = \frac{\sum_k (Q_{ik} - \bar{Q}_i)(Q_{jk} - \bar{Q}_j)}{\left\{ \sum_k (Q_{ik} - \bar{Q}_i)^2 \sum_k (Q_{jk} - \bar{Q}_j)^2 \right\}^{1/2}}, \quad (2.27)$$

donde

$$\bar{Q}_i = n_{ij}^{-1} \sum_k Q_{ik} \quad (2.28)$$

son las sumas sobre k extendidas a todos los puntos temporales en los que tanto el sitio i como el j tienen datos, y n_{ij} es el número de dichos puntos. El valor de correlación medio se define como:

$$\bar{r} = \left\{ \frac{1}{2} N(N-1) \right\}^{-1} \sum_{i \leq j} \sum_{j \leq N} r_{ij}. \quad (2.29)$$

Cabe destacar que son posibles otras alternativas para incluir la dependencia en la simulación. Sin embargo, según Hosking y Wallis (1997), su principal efecto - el aumento del error cuadrático medio de los cuantiles y la curva regional de frecuencia - queda suficientemente valorado con la matriz de equicorrelación (ecuación 2.26).

Con el proceso de simulación, las estimaciones de los cuantiles se calculan para ciertas probabilidades de no excedencia. En la repetición m -ésima de la estación i , denótese la estimación del cuantil de probabilidad de no excedencia F como $\hat{Q}_i^{[m]}(F)$. Su error relativo es $\{\hat{Q}_i^{[m]}(F) - Q_i(F)\} / Q_i(F)$, y esta cantidad puede ser cuadrada y promediada sobre las M repeticiones para obtener el error cuadrático medio relativo de los estimadores. Para valores elevados de M , dicho error se expresa como:

$$R_i(F) = \left[M^{-1} \sum_{m=1}^M \left\{ \frac{\hat{Q}_i^{[m]}(F) - Q_i(F)}{Q_i(F)} \right\}^2 \right]^{1/2}. \quad (2.30)$$

El resumen de la precisión de las estimaciones de los cuantiles a lo largo de todas las estaciones de la región viene dado por el valor medio regional del error cuadrático medio:

$$R^R(F) = N^{-1} \sum_{i=1}^N R_i(F). \quad (2.31)$$

Adicionalmente, se pueden calcular valores análogos para la estimación de la curva de frecuencia. Sea $q_i(F)$ la curva del sitio i , definida por:

$$Q_i(F) = \mu_i q_i(F). \quad (2.32)$$

Las cantidades $q_i(F)$ se necesitan para la simulación de regiones heterogéneas, mientras que en las homogéneas cada $q_i(F)$ es igual a la curva regional de frecuencia $q(F)$. En la repetición m -ésima, denótese la curva regional estimada como $\hat{q}^{[m]}(F)$. Las medidas de precisión de las estimaciones de las curvas se definen también con las ecuaciones 2.30 y 2.31, reemplazando $Q_i(F)$ y $\hat{Q}_i^{[m]}(F)$ por $q_i(F)$ y $\hat{q}^{[m]}(F)$, respectivamente. Estas medidas son de especial interés para aquellas situaciones en las que únicamente interesa la estimación de las curvas de frecuencia, como ocurre cuando el índice de avenida se estima mediante métodos que no implican el empleo de datos locales.

Otros medidores útiles, particularmente cuando la distribución de las estimaciones es asimétrica, son los cuantiles empíricos de dicha distribución. Estos pueden ser obtenidos

calculando el ratio entre los valores estimados y los verdaderos - $\hat{Q}_i(F)/Q_i(F)$ en el caso de los cuantiles, y $\hat{q}(F)/q_i(F)$ en las curvas – promediando dichos valores para todas las estaciones de la región y acumulando sobre las diferentes realizaciones un histograma de los valores que toma el ratio. Por ejemplo, para una probabilidad de no excedencia F puede determinarse que el 5% de los valores simulados de $\hat{Q}(F)/Q(F)$ se encuentra por debajo del valor $L_{.05}(F)$, mientras que otro 5% queda por encima de $U_{.05}(F)$; así, el 90% de la distribución de $\hat{Q}(F)/Q(F)$ se incluye en el intervalo:

$$L_{.05}(F) \leq \frac{\hat{Q}(F)}{Q(F)} \leq U_{.05}(F), \quad (2.33)$$

e invirtiéndolo para expresar Q en términos de \hat{Q} :

$$\frac{\hat{Q}(F)}{U_{.05}(F)} \leq Q(F) \leq \frac{\hat{Q}(F)}{L_{.05}(F)}. \quad (2.34)$$

La expresión 2.34 adquiere forma de intervalo de confianza, pero sólo puede ser interpretada como tal si la distribución de $\hat{Q}(F)/Q(F)$ es independiente de los parámetros implicados en la especificación del modelo que subyace en el método del índice de avenida; para el algoritmo regional de L -momentos, las medias locales de las observaciones y los ratios medios regionales de L -momentos. Pese a que en la práctica esta independencia no se cumple, el intervalo constituye una buena indicación de la variación entre las cantidades estimadas y las verdaderas.

Los límites $\hat{Q}(F)/U_{.05}(F)$ y $\hat{Q}(F)/L_{.05}(F)$ reciben el nombre de “límites de error al 90%” para $\hat{Q}(F)$, y pueden ser de gran utilidad en la cola inferior de la distribución. Si ésta toma valores negativos, puede ocurrir que $L_{.05}(F)$ sea muy pequeño o negativo, conduciendo a un límite superior demasiado elevado o incluso infinito. En estos casos, el valor medio regional del error cuadrático medio $R^R(F)$ de las estimaciones (ecuación 2.31) constituye una medida de precisión más eficiente.

Hay que destacar, no obstante, que esta forma de simulación basada en los límites de la ecuación 2.34 es menos exacta que la construcción formal de intervalos de confianza, pero aun así aporta estimaciones razonables de los errores que se pueden esperar en los cuantiles y las curvas de frecuencia. En este sentido, la precisión viene condicionada por el número de repeticiones M de la simulación. Valores de $M = 100$ pueden resultar suficientes, aunque Hosking y Wallis (1997) recomiendan $M = 1000$ o incluso $M = 10000$ de cara a un mayor rigor.

2.4.- EL ENTORNO DE PROGRAMACIÓN R

La metodología descrita del análisis regional basado en el empleo conjunto del índice de avenida y los L -momentos se implementa mediante R.

R es un lenguaje y entorno de programación para análisis estadístico y gráfico. Se trata de un proyecto de software libre (<http://www.r-project.org/>), resultado de la implementación del premiado lenguaje S. Iniciado por el Departamento de Estadística de la Universidad de Auckland en 1993, su desarrollo actual es responsabilidad del *R Development Core Team*.

R y S-Plus - versión comercial de S - son, probablemente, los dos lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy populares en el campo de la investigación biomédica, la bioinformática y las matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con finalidades específicas de cálculo y representación gráfica.

R engloba un amplio abanico de herramientas estadísticas (modelos lineales y no lineales, tests estadísticos, análisis de series temporales, algoritmos de clasificación y agrupamiento, etc.) y gráficas. Al igual que S, se trata de un lenguaje de programación que permite a los usuarios la extensión y definición de sus propias funciones. De hecho, gran parte de las funciones de R están escritas en el mismo R, aunque para algoritmos computacionalmente exigentes es posible desarrollar bibliotecas en C, C++ o Fortran que se cargan dinámicamente.

R hereda de S su orientación a objetos. Tiene capacidad para integrarse con distintas bases de datos y existen bibliotecas que facilitan su utilización desde lenguajes de programación interpretados, como Perl y Python. Incluso puede emplearse como herramienta de cálculo numérico, campo en el que puede ser tan eficaz como otras herramientas específicas tipo GNU Octave y su versión comercial, MATLAB.

Además, el punto fuerte de R reside en que forma parte de un proyecto colaborativo y abierto. Cualquier usuario puede publicar paquetes que extienden su configuración básica, y para ello existe un repositorio oficial que los ordena según su naturaleza y función. A modo de ejemplo, existen grupos de paquetes relacionados con estadística bayesiana, econometría y series temporales, entre muchos otros.

Así, todo el procedimiento de análisis regional descrito se codifica y ejecuta con la ayuda de los paquetes *lmom* (Hosking, 2009) y *lmomRFA* (Hosking, 2010). El código elaborado para el estudio de la cuenca del Arga se recoge en el anexo II.

3.- RESULTADOS Y DISCUSIÓN

3.1.- Análisis regional de frecuencias

3.1.1.- Filtrado primario de las observaciones

3.1.2.- Identificación de regiones homogéneas

3.1.3.- Elección de la distribución de frecuencia

3.1.4.- Estimación de los cuantiles de frecuencia

3.2.- Comparación con los métodos de análisis local

3.1.- ANÁLISIS REGIONAL DE FRECUENCIAS

3.1.1.- FILTRADO PRIMARIO DE LAS OBSERVACIONES. DISCORDANCIA.

Se parte de las series de observaciones de precipitación máxima en 24 h. registradas en las estaciones indicadas en el apartado 2.2 (tabla 2.1). Antes de comenzar con el análisis propiamente dicho, se procedió a su ordenamiento y limpieza. En primer lugar, se creó una aplicación en MATLAB (software matemático de cálculo numérico orientado a matrices) que permitió juntar los diferentes archivos anuales “.txt” existentes por estación en uno solo, consiguiéndose la ordenación cronológica de las observaciones de cada lugar. Además de devolver un único archivo por estación, la aplicación se programó para calcular automáticamente los valores diarios máximos anuales y detectar todos aquellos días que carecían de registro y figuraban en blanco.

Se comprobó cómo en algunas estaciones, sobretudo en las de mayor registro, faltaban datos en días sueltos, en meses enteros o incluso en años completos. La presencia de valores erróneos, aunque con menos frecuencia, también estuvo presente. Todo ello fue depurado. Se suprimieron aquellos años en los que se consideró significativa la ausencia de ciertos registros diarios, fuera porque estos no se encontraban disponibles para buena parte del año, por su carencia precisamente durante la estación lluviosa o por comparación de los días sin registro de cada estación con los valores de las estaciones colindantes.

Con las series así depuradas, de longitud muestral indicada en la tabla 3.1, se obtienen los *L*-momentos cuyos valores también se presentan en la tabla.

Tabla 3.1. *Valores de los ratios de L-momentos de las estaciones.*

Estación	longitud, n	media, l_1	<i>L</i> -CV, t	<i>L</i> -asimetría, t_3	<i>L</i> -kurtosis, t_4
Belzunce	20	58.55	0.1974	0.2260	0.1889
Erro	34	63.00	0.1539	0.1442	0.1428
Espinal	37	62.31	0.1945	0.1317	0.1044
Eugui	40	68.03	0.1502	0.2185	0.1520
Ilundáin	28	49.84	0.1600	0.2046	0.0325
Iraizoz	31	62.31	0.1325	0.2630	0.0980
Iroz	27	46.37	0.1725	0.0797	0.1352
Monreal	27	49.27	0.2030	0.2462	0.1675
Noáin	34	46.19	0.2192	0.2981	0.2208
Olagüe	32	56.79	0.1586	0.1585	0.1108
Otazu	33	46.92	0.1601	0.1822	0.0871
Pamplona	116	52.96	0.2227	0.2562	0.1405
Velate	21	67.08	0.1306	0.0306	-0.0390
Zuazu	21	50.40	0.2017	0.2888	0.2288
Zubiri	36	56.10	0.1434	0.1201	0.1858

Se comienza con una inspección preliminar de todas las estaciones, representando gráficamente sus valores L -CV y L -kurtosis frente a L -asimetría con el objetivo de obtener una primera aproximación:

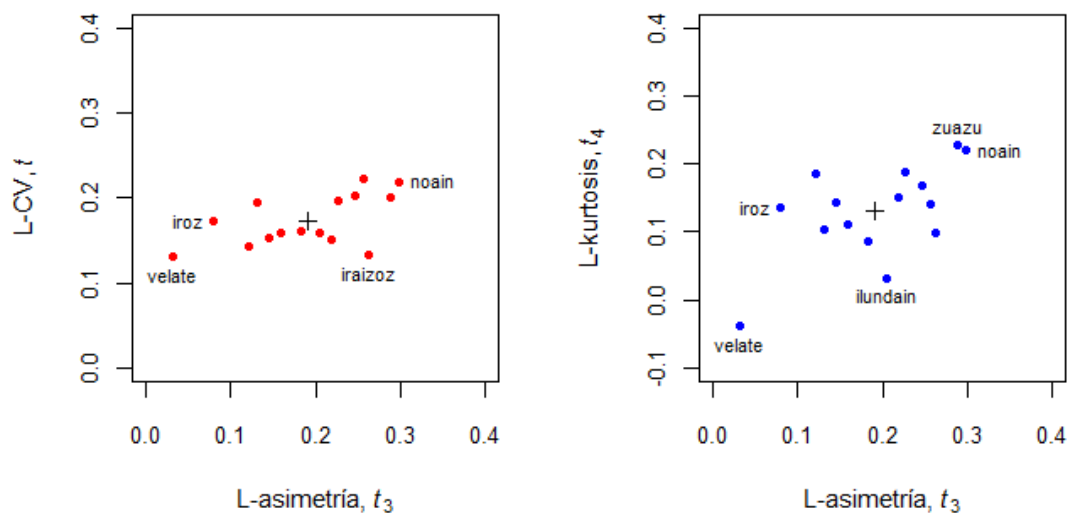


Fig. 3.1. Representación gráfica de los ratios de las estaciones.

A primera vista, parece que Iraizoz, Iroz y sobretodo Velate pueden ser discordantes con el resto de estaciones. Sus valores quedan bastante alejados de la media del grupo, representada por el símbolo +.

El cálculo de las medidas de discordancia D_i según la ecuación 2.3 arroja los siguientes resultados:

Tabla 3.2. Valores de la discordancia D_i de las estaciones.

Estación	D_i	Estación	D_i
Belzunce	0.31	Noáin	0.98
Erro	0.44	Olagüe	0.09
Espinal	0.92	Otazu	0.19
Eugui	0.66	Pamplona	1.33
Ilundáin	1.27	Velate	2.32
Iraizoz	2.20	Zuazu	0.81
Iroz	1.27	Zubiri	1.86
Monreal	0.36		

Atendiendo a la tabla 2.2, el umbral crítico de la medida de discordancia para un grupo de 15 ó más estaciones es 3. Ni siquiera Velate, con un máximo de 2.32, alcanza el valor crítico.

Por tanto, se puede concluir que ninguna de las estaciones es discordante con el resto. Se puede considerar que ninguna serie requiere entonces de un análisis más profundo y exhaustivo para su filtrado.

3.1.2.- IDENTIFICACIÓN DE REGIONES HOMOGÉNEAS

Realizado el filtrado primario de las observaciones, se procede a la identificación de regiones homogéneas.

Se lleva a cabo un análisis cluster de las estaciones, definiéndose mediante cuatro variables físicas que se consideran relacionadas con la precipitación máxima en 24 h.: precipitación media anual, altitud, latitud y temperatura media anual.

Tabla 3.3. *Descriptor físicos de las estaciones* (Fuente: Estudio Agroclimático de Navarra – Departamento de Agricultura, Ganadería y Alimentación – Gobierno de Navarra, 2001).

Estación	Precipitación media anual (mm)	Altitud (m)	Latitud (°N)	Temperatura media anual (°C)
Belzunce	1100.0	530	42.8983	11.7
Erro	1310.3	671	42.9424	9.8
Espinal	1521.1	870	42.9789	8.8
Eugui	1535.8	616	42.9786	10.3
Ilundáin	837.1	572	42.7757	12.1
Iraizoz	1321.8	550	42.9933	10.8
Iroz	890.8	478	42.8612	12.4
Monreal	783.1	533	42.7046	13.2
Noáin	714.8	456	42.7605	12.4
Olagüe	1246.5	549	42.9635	11.1
Otazu	807.1	387	42.7849	12.5
Pamplona	813.4	453	42.8180	12.5
Velate	1450.0	797	43.0490	10.5
Zuazu	825.9	627	42.7367	12.6
Zubiri	1225.0	526	42.9304	11.5

Cada una de estas variables se estandariza dividiéndola por su desviación típica. El optar por esta forma de estandarización se explica por el hecho de que, normalmente, los algoritmos de cluster basan el agrupamiento en las distancias euclídeas entre puntos, siendo precisamente este cálculo el implicado en la desviación estándar. Con ella, además, se consigue un mismo escenario de dispersión para todas las variables que después pueden ser ponderadas. De ahí que estandarizar dividiendo por el máximo de la variable – con el objetivo de obtener valores comprendidos entre 0 y 1 - o por su rango no resulte recomendable para el estudio, pudiendo incluso inducir a una descompensación en el peso de las dispersiones.

Por otro lado, se pondera cada variable en función de su peso considerado en la regionalización. Se asigna un coeficiente de 1.5 a la precipitación media anual y a la altitud por estimarse como las más relevantes. A la latitud se le otorga un coeficiente de 1, y a la temperatura media anual, asumida como menos importante, se le asigna un valor de 0.5:

Tabla 3.4. Transformación de las características físicas.

Variable física, X	Variable cluster, Y
Precipitación media anual (mm)	$Y = 1.5 \frac{X}{s(X)}$
Altitud (m)	$Y = 1.5 \frac{X}{s(X)}$
Latitud (°N)	$Y = 1 \frac{X}{s(X)}$
Temperatura media anual (°C)	$Y = 0.5 \frac{X}{s(X)}$

El algoritmo empleado en el análisis cluster es el método Ward, procedimiento jerárquico aglomerativo que agrupa las estaciones de manera que la varianza intragrupal de la estructura formada sea mínima. De forma más detallada: la pérdida de información que se produce al integrar las distintas estaciones en clusters puede medirse a través de la suma total de los cuadrados de las desviaciones entre cada estación y la media del cluster en el que se integra. En cada paso del análisis, por tanto, el algoritmo considera la posibilidad de la unión de cada par de grupos y opta por la fusión de aquellos que menos incrementen la suma de los cuadrados de las desviaciones al unirse, dando como resultado agrupaciones que tienden a ser equilibradas en tamaño.

Se ejecuta el análisis cluster y se comprueba mediante el cálculo de la medida de heterogeneidad H planteada en la ecuación 2.9:

$$H = \frac{(V - \mu_V)}{\sigma_V}$$

donde V es la desviación estándar ponderada de los L -CV locales (ecuación 2.8), y μ_V y σ_V son, respectivamente, la media y la desviación estándar de los valores de V procedentes de las simulaciones de una región con el mismo número de estaciones y las mismas longitudes muestrales que la real. Atendiendo a la recomendación de Hosking y Wallis (1997), se realizan 500 simulaciones de la distribución Kappa por ser tertraparamétrica e incluir entre sus casos especiales buena parte de las distribuciones empleadas en estudios medioambientales. En la tabla 3.5 se muestran los resultados:

Tabla 3.5. Resultado del análisis cluster y medida de la heterogeneidad H .

Cluster	Tamaño	Estaciones	Heterogeneidad H
Norte	8	Belzunce, Erro, Espinal, Eugui, Iraizoz, Olagüe, Velate y Zubiri.	1.02
Sur	7	Ilundáin, Iroz, Monreal, Noáin, Otazu, Pamplona y Zuazu.	0.96

Nota: el resultado del análisis cluster es el mismo aun tanteándose diferentes formas de estandarización y ponderación de las variables, e incluso en ausencia de ellas.

Obsérvese la diferenciación de dos regiones, a las que se nombra como Norte y Sur dada la ubicación de sus estaciones en la cuenca, y de tamaño 8 y 7 respectivamente. La región Norte, con un valor de $H = 1.02$, resulta “posiblemente heterogénea” por estar comprendido entre 1 y 2. La región Sur, con valor de $H = 0.96$, resulta “aceptablemente homogénea” al ser inferior a 1.

Sin embargo, se toma en cuenta una consideración subjetiva. El conocimiento físico de la cuenca del Arga lleva a valorar la posibilidad de trasladar la estación de Iroz del cluster Sur al cluster Norte. Observando la figura 2.4 y atendiendo a la ubicación de dicha estación – contrariamente al resto de estaciones de la región Sur, queda alejada de la cuenca de Pamplona y en zona orográficamente más montañosa – parece probable que la zona Norte sea el cluster más adecuado para enmarcarla. La representación de sus L -momentos con el grupo Sur confirma esta posibilidad, pues sus valores quedan bastante alejados de los del resto de estaciones y de la media:

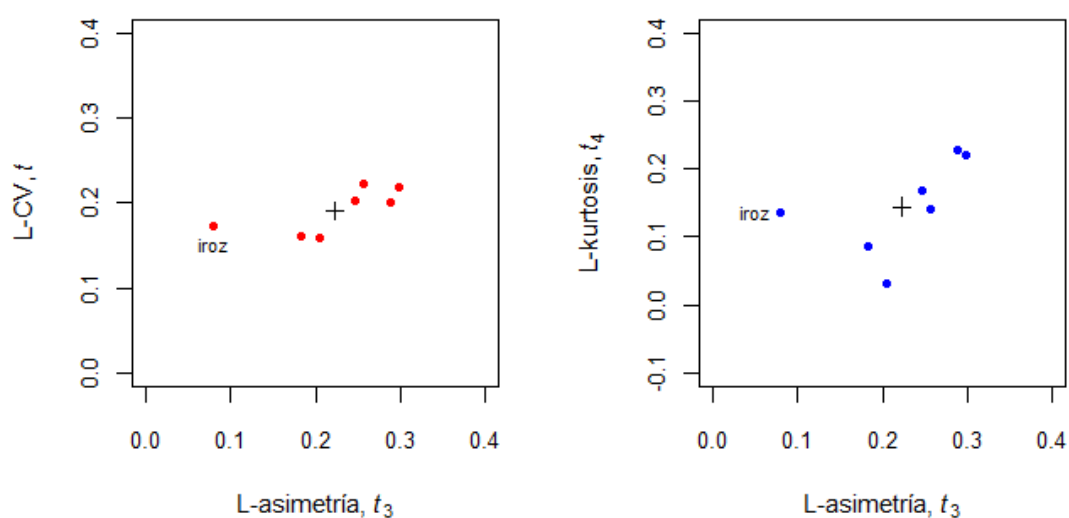


Fig. 3.2. Representación gráfica de Iroz en la región Sur.

Se traslada entonces la estación de Iroz a la región Norte y se recalculan los valores de heterogeneidad H , obteniéndose los siguientes resultados:

Tabla 3.6. Resultados ajustados de la medida de heterogeneidad H .

Cluster	Tamaño	Estaciones	Heterogeneidad H
Norte	9	Belzunce, Erro, Espinal, Eugui, Iraizoz, Iroz, Olagüe, Velate y Zubiri.	0.76
Sur	6	Ilundáin, Monreal, Noáin, Otazu, Pamplona y Zuazu.	0.92

Apréciase cómo, efectivamente, con el traslado de Iroz se consigue convertir ambas regiones en “aceptablemente homogéneas”, al arrojar valores de H de 0.76 y 0.92, inferiores a 1.

Queda pues establecida la regionalización definitiva (tabla 3.6).

Con el fin de completar el análisis, se realiza una nueva medida de discordancia D_i para cada estación dentro de su región asignada:

Tabla 3.7. Valores de la discordancia D_i de las estaciones en su región.

Norte	D_i	Sur	D_i
Belzunce	1.17	Ilundáin	1.61
Erro	0.17	Monreal	0.30
Espinal	1.14	Noáin	0.54
Eugui	0.41	Otazu	1.33
Iraizoz	1.68	Pamplona	1.26
Iroz	0.68	Zuazu	0.95
Olagüe	0.02		
Velate	2.16		
Zubiri	1.56		

Atendiendo a la tabla 2.2, el umbral crítico de la medida de discordancia para un grupo de 9 estaciones es 2.329, y para un grupo de 6 estaciones, 1.648. No existe caso en que se superen esos valores, por lo que ninguna estación es discordante con las de su región.

Con todo, se concluye habiendo identificado dos regiones homogéneas en la cuenca del Arga: la Norte, con 9 estaciones, y la Sur, con 6 estaciones.

3.1.3.- ELECCIÓN DE LA DISTRIBUCIÓN DE FRECUENCIA

Definidas las regiones, se procede a determinar la función de distribución más apropiada para cada una de ellas.

Antes de calcular las medidas de bondad de ajuste, se toma en cuenta una consideración importante. Los valores de heterogeneidad H se han estimado asumiendo el cumplimiento de las suposiciones de las que parte el método del índice de avenida. Sin embargo, atendiendo a la superficie de la cuenca y a la proximidad entre ciertas estaciones, parece que la independencia entre series no está garantizada, sino que cabe esperar la existencia de cierta correlación. La inspección visual de los valores diarios anuales y la detección de coincidencias en las fechas de algunos eventos máximos en diferentes sitios así lo corroboran.

Por ello, se implementa en el cálculo una matriz de correlaciones para cada región que tiene en cuenta los efectos de la dependencia entre estaciones. Si bien escoger un mismo valor de equicorrelación (ecuación 2.26) puede resultar suficiente, se opta por comparar los años de registro en común de cada par de estaciones para obtener los diferentes coeficientes y lograr así una mayor precisión.

Tabla 3.8. Matriz de correlaciones de la región Norte.

	Belzunce	Erro	Espinal	Eugui	Iraizoz	Iroz	Olagüe	Velate	Zubiri
Belzunce	1	0.689	0.233	0.813	0.570	0.274	0.690	0.582	0.736
Erro	0.689	1	0.267	0.621	0.327	0.681	0.581	0.415	0.638
Espinal	0.233	0.267	1	0.455	0.256	0.412	0.205	0.212	0.214
Eugui	0.813	0.621	0.455	1	0.692	0.338	0.609	0.449	0.666
Iraizoz	0.570	0.327	0.256	0.692	1	0.168	0.605	0.151	0.369
Iroz	0.274	0.681	0.412	0.338	0.168	1	0.335	0.414	0.530
Olagüe	0.690	0.581	0.205	0.609	0.605	0.335	1	0.209	0.560
Velate	0.582	0.415	0.212	0.449	0.151	0.414	0.209	1	0.291
Zubiri	0.736	0.638	0.214	0.666	0.369	0.530	0.560	0.291	1

Tabla 3.9. Matriz de correlaciones de la región Sur.

	Ilundáin	Monreal	Noáin	Otazu	Pamplona	Zuazu
Ilundáin	1	0.794	0.708	0.280	0.644	0.505
Monreal	0.794	1	0.622	0.281	0.447	0.668
Noáin	0.708	0.622	1	0.430	0.752	0.505
Otazu	0.280	0.281	0.430	1	0.520	0.262
Pamplona	0.644	0.447	0.752	0.520	1	0.474
Zuazu	0.505	0.668	0.505	0.262	0.474	1

Se recalculan nuevamente los valores de heterogeneidad H de la ecuación 2.9, empleándose de nuevo 500 simulaciones de la distribución Kappa y considerando la correlación entre las estaciones:

Tabla 3.10. Resultados de la medida de heterogeneidad H considerando la dependencia entre estaciones.

Región	Heterogeneidad H
Norte	1.08
Sur	1.09

Apréciase el aumento de la medida de heterogeneidad en ambas regiones con respecto a la tabla 3.6. El cálculo arroja un valor de 1.08 para la Norte y 1.09 para la Sur. Se califican ahora como “posiblemente heterogéneas”, al superar el límite 1. No obstante, la cercanía a dicho límite permite considerar que, en el peor de los casos, se trata de una situación de heterogeneidad moderada. El hecho de no alcanzar el umbral 2 - a partir del cual las regiones serían “definitivamente heterogéneas” - confirma la posibilidad de conseguir estimaciones precisas.

Se procede a valorar, para cada región, el ajuste de las siguientes funciones tripamétricas recomendadas por Hosking y Wallis (1997) por incluir entre sus casos particulares las distribuciones más empleadas en el estudio de variables ambientales:

- Logística Generalizada (GLO). Parámetros (3): ξ (localización), α (escala), k (forma).

Su función de distribución se expresa como:

$$F(x) = \frac{1}{(1 + e^{-y})} \quad \text{siendo } y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases}$$

- General de Valores Extremos (GEV). Parámetros (3): ξ (localización), α (escala), k (forma).

Su función de distribución se expresa como:

$$F(x) = e^{-e^{-y}} \quad \text{siendo } y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases}$$

- General de Pareto (GPA). Parámetros (3): ξ (localización), α (escala), k (forma).

Su función de distribución se expresa como:

$$F(x) = 1 - e^{-y} \quad \text{siendo } y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases}$$

- LogNormal (LN3). Parámetros (3): ξ (localización), α (escala), k (forma).

Su función de densidad se expresa como:

$$f(x) = \frac{e^{ky - y^2/2}}{\alpha\sqrt{2\pi}} \quad \text{siendo } y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases}$$

- Pearson tipo III (PE3). Parámetros (3): μ (localización), σ (escala), γ (forma).

Siendo $\alpha = 4/\gamma^2$, $\beta = \frac{1}{2}\sigma|\gamma|$, $\xi = \mu - 2\sigma/\gamma$, su función de distribución se expresa como:

$$F(x) = \begin{cases} G\left(\alpha, \frac{x-\xi}{\beta}\right) / \Gamma(\alpha), & \gamma > 0 \\ 1 - G\left(\alpha, \frac{\xi-x}{\beta}\right) / \Gamma(\alpha), & \gamma < 0 \end{cases}$$

siendo $\Gamma(\cdot)$ la función Gamma y $G(\cdot)$ la función Gamma incompleta.

Sus descripciones con mayor detalle se recogen en el anexo I.

3.1.3.1.- REGIÓN NORTE

Se presenta, en primer lugar, un diagrama de L -momentos con las cinco distribuciones de cara a obtener una aproximación inicial:

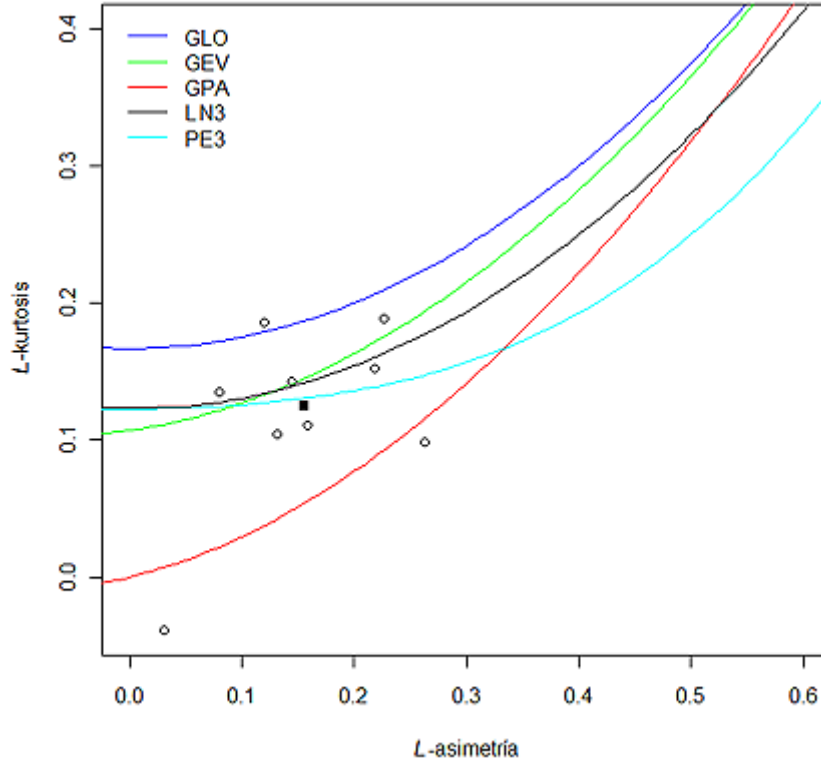


Fig. 3.3. Diagrama de L -momentos de las distribuciones en la región Norte.

En la figura 3.3 se reflejan en trazos las relaciones entre los L -asimetría y L -kurtosis poblacionales de cada una de las distribuciones, asumiendo los momentos muestrales como representativos de los poblacionales. Los círculos representan cada una de las estaciones, mientras que el cuadrado negro hace referencia a la media regional ponderada. Obsérvese cómo esta última parece aproximarse a la distribución Pearson tipo III y, en menor medida, a la General de Valores Extremos y a la LogNormal. Las distribuciones Logística Generalizada y General de Pareto quedan más alejadas.

Se calcula para cada distribución la medida de bondad de ajuste Z de la ecuación 2.21:

$$Z^{DIST} = (\tau_4^{DIST} - t_4^R + B_4) / \sigma_4$$

donde τ_4^{DIST} es el ratio de L -kurtosis de la distribución ajustada a los valores regionales de orden inferior, t_4^R el ratio medio regional de L -kurtosis, y B_4 y σ_4 son, respectivamente, el sesgo y la desviación estándar de los valores de t_4^R procedentes de las simulaciones de una región con el mismo número de estaciones y las mismas longitudes muestrales que la real. Siguiendo la recomendación de Hosking y Wallis

(1997), se aprovechan las 500 simulaciones de la distribución Kappa empleadas en el cálculo de la heterogeneidad H . En la tabla 3.11 se recogen los resultados:

Tabla 3.11. Valores del estadístico Z para las diferentes distribuciones en la región Norte.

Z^{GLO}	Z^{GEV}	Z^{GPA}	Z^{LN3}	Z^{PE3}
2.93	1.01	-3.16	0.88	0.38

Atendiendo al criterio $|Z| \leq 1.64$ se comprueba cómo, efectivamente, son tres las distribuciones candidatas para la región Norte: General de Valores Extremos, LogNormal y Pearson tipo III. Se descartan la Logística Generalizada y la General de Pareto.

Se representa la curva regional de frecuencia de cada una de las distribuciones candidatas para compararlas:

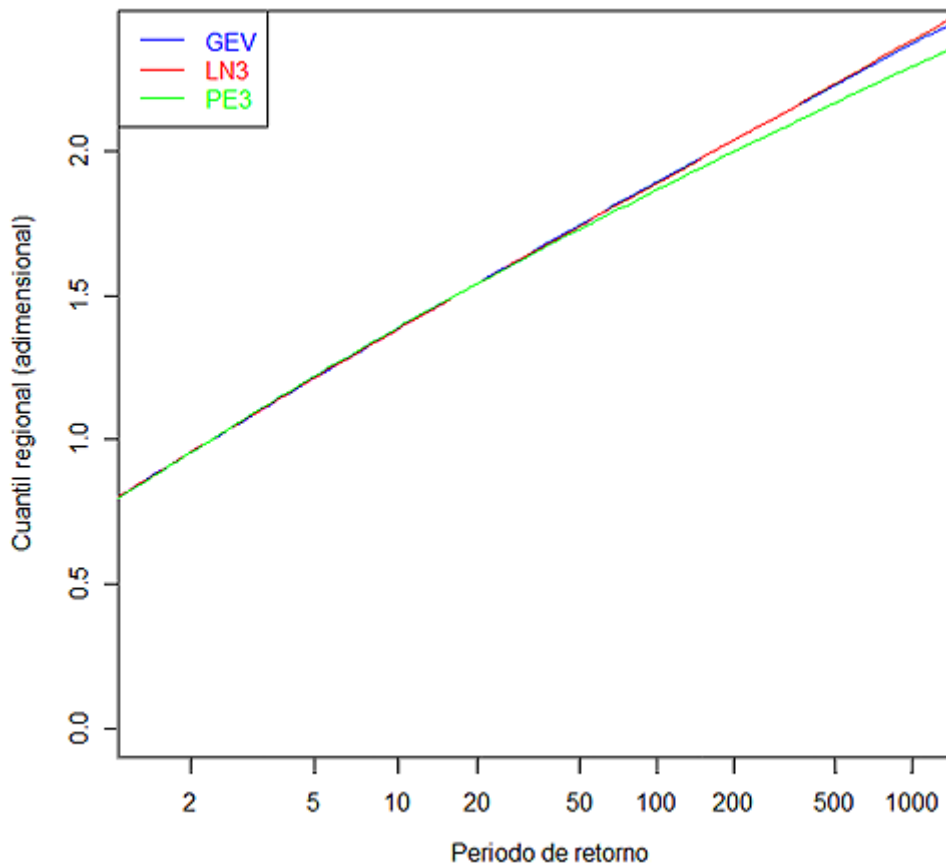


Fig. 3.4. Curva regional de frecuencia de las distribuciones candidatas de la región Norte.

En ellas se expresan los cuantiles regionales adimensionalizados – obtenidos a partir de todas las observaciones de la región – frente a diferentes periodos de retorno habituales. Apréciense cómo las curvas son prácticamente similares. Su forma es casi la misma hasta el periodo de retorno de 50 años, a partir del cual surgen pequeñas diferencias lógicas de tratar con los cuantiles extremos de la cola de la distribución. Estas diferencias son más notables en la distribución Pearson tipo III frente a la General de Valores Extremos y la LogNormal.

Dada la similitud, a efectos prácticos, la elección de una u otra distribución puede resultar indiferente. Sin embargo, es recomendable la opción por el modelo más robusto, que no implica necesariamente el mejor ajuste dado que las series evaluadas constituyen tan sólo una muestra de las múltiples posibles de la población. Se calcula entonces el error cuadrático medio relativo expresado en la ecuación 2.30:

$$R_i(F) = \left[M^{-1} \sum_{m=1}^M \left\{ \frac{\hat{Q}_i^{[m]}(F) - Q_i(F)}{Q_i(F)} \right\}^2 \right]^{1/2}$$

donde $\hat{Q}_i^{[m]}(F)$ es la estimación del cuantil de probabilidad de no excedencia F en la m -ésima simulación de la estación i , y M indica el número de simulaciones de una región con el mismo número de estaciones y las mismas longitudes muestrales que la real. Atendiendo a la recomendación de Hosking y Wallis (1997), se ejecutan $M = 1000$ simulaciones de cada una de las distribuciones para los cuantiles asociados a los periodos de retorno T de interés:

Tabla 3.12. Error cuadrático medio relativo de las distribuciones candidatas en la región Norte.

T (años)	ECM relativo		
	GEV	LN3	PE3
2	0.0169	0.0177	0.0180
5	0.0190	0.0187	0.0187
10	0.0301	0.0307	0.0313
20	0.0438	0.0443	0.0438
50	0.0663	0.0636	0.0592
100	0.0861	0.0788	0.0699
200	0.1074	0.0941	0.0798
500	0.1372	0.1142	0.0918
1000	0.1608	0.1293	0.1002

Se observa cómo, salvo para los periodos de retorno de 2, 5 y 10 años, la distribución Pearson tipo III arroja el menor error. Para los cuantiles asociados a los periodos de

retorno más extremos, sus valores de error son considerablemente inferiores a los de las distribuciones General de Valores Extremos y LogNormal.

Por tanto, se concluye que la distribución Pearson tipo III es la más adecuada para caracterizar la región Norte.

3.1.3.2.- REGIÓN SUR

Se ejecuta el mismo procedimiento para la región Sur. Se comienza con el diagrama de L -momentos de las cinco distribuciones para obtener la primera aproximación:

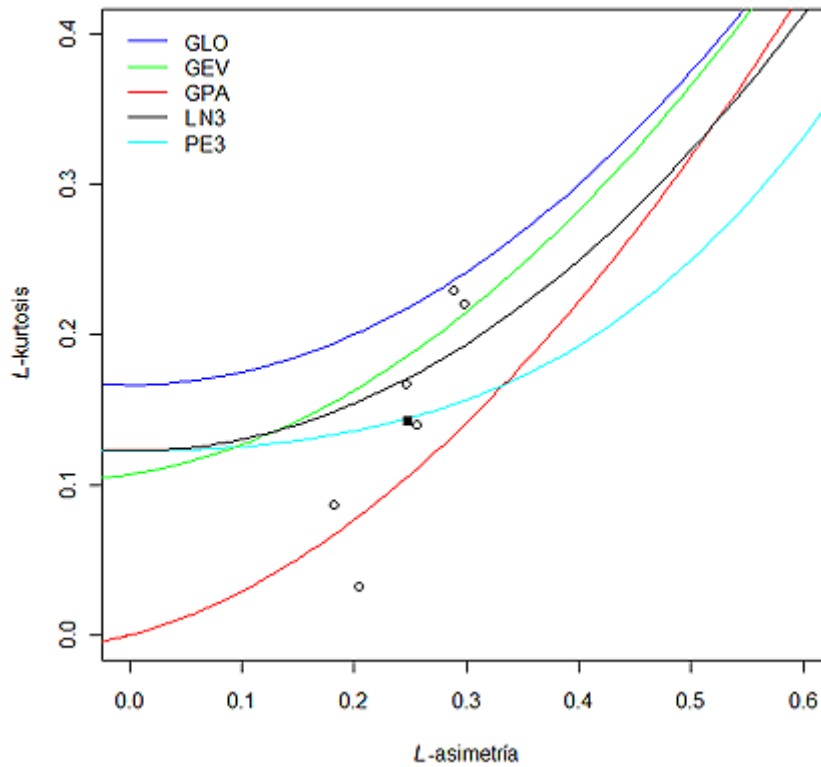


Fig. 3.5. Diagrama de L -momentos de las distribuciones en la región Sur.

De nuevo la figura 3.5 refleja en trazos las relaciones entre los L -asimetría y L -kurtosis poblacionales de cada una de las distribuciones, asumiendo los momentos muestrales como representativos de los poblacionales. En esta ocasión, obsérvese cómo la media regional ponderada, representada por el cuadrado negro, también parece aproximarse a la distribución Pearson tipo III y, en menor medida, a la LogNormal y a la General de Pareto. Las distribuciones Logística Generalizada y General de Valores Extremos quedan ahora más alejadas.

Se calcula nuevamente la medida de bondad de ajuste Z de la ecuación 2.21 para cada distribución:

$$Z^{DIST} = (\tau_4^{DIST} - t_4^R + B_4) / \sigma_4$$

donde τ_4^{DIST} es el ratio de L -kurtosis de la distribución ajustada a los valores regionales de orden inferior, t_4^R el ratio medio regional de L -kurtosis, y B_4 y σ_4 son, respectivamente, el sesgo y la desviación estándar de los valores de t_4^R procedentes de

las simulaciones de una región con el mismo número de estaciones y las mismas longitudes muestrales que la real. Con la recomendación de Hosking y Wallis (1997), se aprovechan las 500 simulaciones de la distribución Kappa empleadas en el cálculo de la heterogeneidad H . En la tabla 3.13 se recogen los resultados:

Tabla 3.13. Valores del estadístico Z para las diferentes distribuciones en la región Sur.

Z^{GLO}	Z^{GEV}	Z^{GPA}	Z^{LN3}	Z^{PE3}
3.46	2.14	-1.12	1.58	0.53

Atendiendo al criterio $|Z| \leq 1.64$ se comprueba cómo, efectivamente, son tres las distribuciones candidatas para la región Sur: General de Pareto, LogNormal y Pearson tipo III. Se descartan la Logística Generalizada y la General de Valores Extremos.

Nuevamente se crea la curva regional de frecuencia de cada una de las distribuciones candidatas para compararlas:

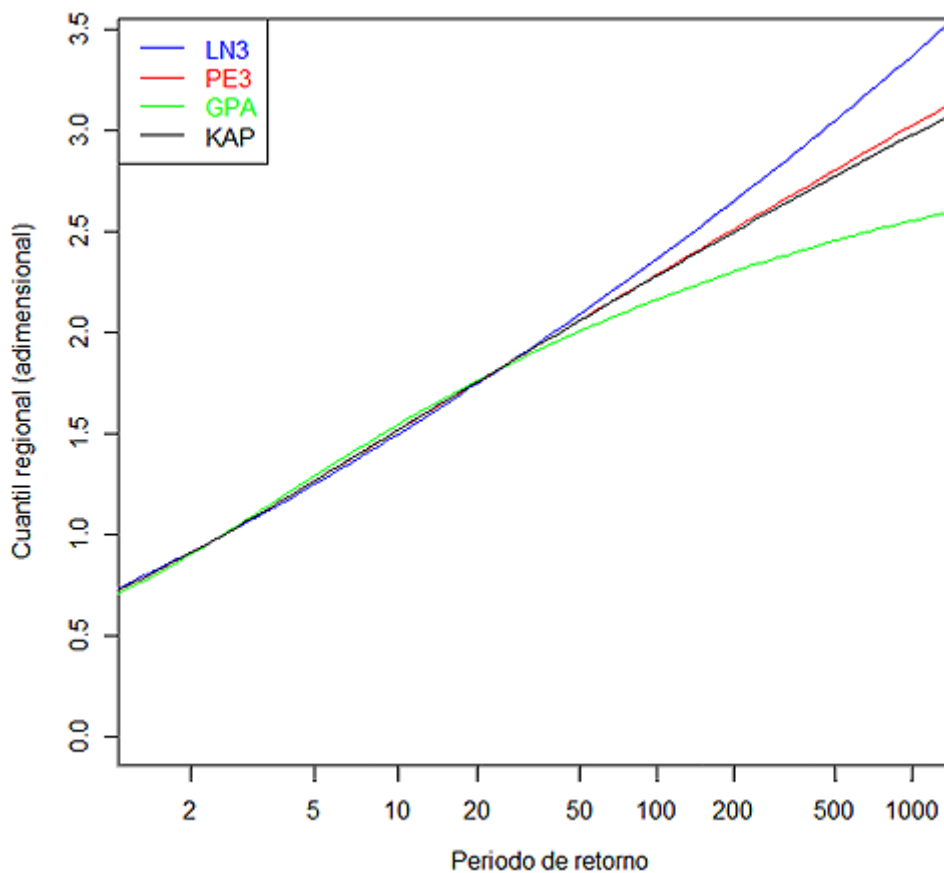


Fig. 3.6. Curva regional de frecuencia de las distribuciones candidatas de la región Sur.

Representando los cuantiles regionales adimensionalizados frente a los diferentes periodos de retorno habituales, apréciase cómo, en este caso, las curvas no son similares. Si bien su forma es parecida hasta el periodo de retorno de 50 años, a partir del mismo surgen diferencias significativas entre las distribuciones, mucho más notables que las que presenta la región Norte, y probablemente debidas al menor número de estaciones y datos disponibles.

La ausencia de similitud complica la elección de la distribución adecuada; de ahí que se opte por valorar también la función tetraparamétrica Kappa (KAP), con vistas a evaluar si alcanza una mayor robustez. De parámetros ξ (localización), α (escala), k y h , su función de distribución se expresa como:

$$F(x) = \left[1 - h \{ 1 - k(x - \xi)/\alpha \}^{1/k} \right]^{1/h}$$

De cara a determinar el modelo más robusto, se calcula de nuevo el error cuadrático medio relativo expresado en la ecuación 2.30:

$$R_i(F) = \left[M^{-1} \sum_{m=1}^M \left\{ \frac{\hat{Q}_i^{[m]}(F) - Q_i(F)}{Q_i(F)} \right\}^2 \right]^{1/2}$$

donde $\hat{Q}_i^{[m]}(F)$ es la estimación del cuantil de probabilidad de no excedencia F en la m -ésima simulación de la estación i , y M indica el número de simulaciones de una región con el mismo número de estaciones y las mismas longitudes muestrales que la real. Con la recomendación de Hosking y Wallis (1997), se ejecutan nuevamente $M = 1000$ simulaciones de cada una de las distribuciones para los cuantiles asociados a los periodos de retorno T de interés:

Tabla 3.14. *Error cuadrático medio relativo de las distribuciones candidatas en la región Sur.*

T (años)	ECM relativo			
	LN3	PE3	GPA	KAP
2	0.0230	0.0240	0.0214	0.0203
5	0.0158	0.0169	0.0175	0.0278
10	0.0324	0.0349	0.0310	0.0352
20	0.0506	0.0506	0.0490	0.0492
50	0.0744	0.0678	0.0745	0.0871
100	0.0919	0.0787	0.0937	0.1252
200	0.1087	0.0881	0.1123	0.1678
500	0.1302	0.0988	0.1355	0.2285
1000	0.1458	0.1059	0.1520	0.2766

Se observa cómo, salvo para los periodos de retorno de 2, 5, 10 y 20 años, la distribución Pearson tipo III arroja el menor error. Para los cuantiles asociados a los periodos de retorno más extremos, sus valores de error son considerablemente inferiores a los de las distribuciones LogNormal y General de Pareto.

Se aprecia también cómo el error de la distribución Kappa se dispara. El emplear un parámetro más en una región de escasas estaciones puede ser la causa de esta incertidumbre tan elevada.

Por tanto, se concluye que, al igual que en la región Norte, la distribución Pearson tipo III es la más adecuada para caracterizar la región Sur.

3.1.4.- ESTIMACIÓN DE LOS CUANTILES DE FRECUENCIA

Seleccionada la distribución de frecuencia más apropiada para cada región, se procede a definir sus parámetros de cara a la estimación de los cuantiles de precipitación en 24 h. de interés.

3.1.4.1.- REGIÓN NORTE

Pearson tipo III (PE3). Parámetros (3): μ (localización), σ (escala), γ (forma).

Siendo $\alpha = 4/\gamma^2$, $\beta = \frac{1}{2}\sigma|\gamma|$, $\xi = \mu - 2\sigma/\gamma$, su función de distribución se expresa como:

$$F(x) = \begin{cases} G\left(\alpha, \frac{x-\xi}{\beta}\right) / \Gamma(\alpha), & \gamma > 0 \\ 1 - G\left(\alpha, \frac{\xi-x}{\beta}\right) / \Gamma(\alpha), & \gamma < 0 \end{cases}$$

siendo $\Gamma(\cdot)$ la función Gamma y $G(\cdot)$ la función Gamma incompleta.

Se comienza por estimar los parámetros de la curva regional de frecuencia de la distribución, igualando los ratios medios regionales de L -momentos de las observaciones a los de la distribución:

Tabla 3.15. *Parámetros de la curva regional de frecuencia de la distribución Pearson tipo III de la región Norte.*

μ	σ	γ
1	0.2893	0.9486

Se computan los cuantiles regionales adimensionales de precipitación en 24 h. para los periodos de retorno T de 2, 5, 10, 20, 50, 100, 200, 500 y 1000 años. Se evalúa asimismo la incertidumbre de su estimación mediante el cálculo del error cuadrático medio absoluto y los límites de error al 90%, establecidos en la ecuación 2.34:

$$\frac{\hat{Q}(F)}{U_{.05}(F)} \leq Q(F) \leq \frac{\hat{Q}(F)}{L_{.05}(F)}$$

Para ello, atendiendo a la recomendación de Hosking y Wallis (1997), se emplean 1000 simulaciones de una región con número de estaciones y longitudes muestrales similares a las de la observada, y con distribución de frecuencia Pearson tipo III, pero cuyos ratios de L -momentos locales son definidos arbitrariamente en un rango menor que el observado para aproximar a los valores poblacionales:

Tabla 3.16. Resultados de las estimaciones regionales de la región Norte.

T (años)	Cuantil regional	ECM	Límite 0.05	Límite 0.95
2	0.955	0.0172	0.922	0.991
5	1.221	0.0229	1.179	1.263
10	1.388	0.0434	1.302	1.478
20	1.540	0.0674	1.400	1.686
50	1.729	0.1024	1.521	1.947
100	1.865	0.1303	1.602	2.140
200	1.996	0.1593	1.680	2.330
500	2.165	0.1987	1.777	2.580
1000	2.290	0.2293	1.845	2.767

Se representan gráficamente los resultados de las estimaciones, indicándose en trazo discontinuo los límites de error:

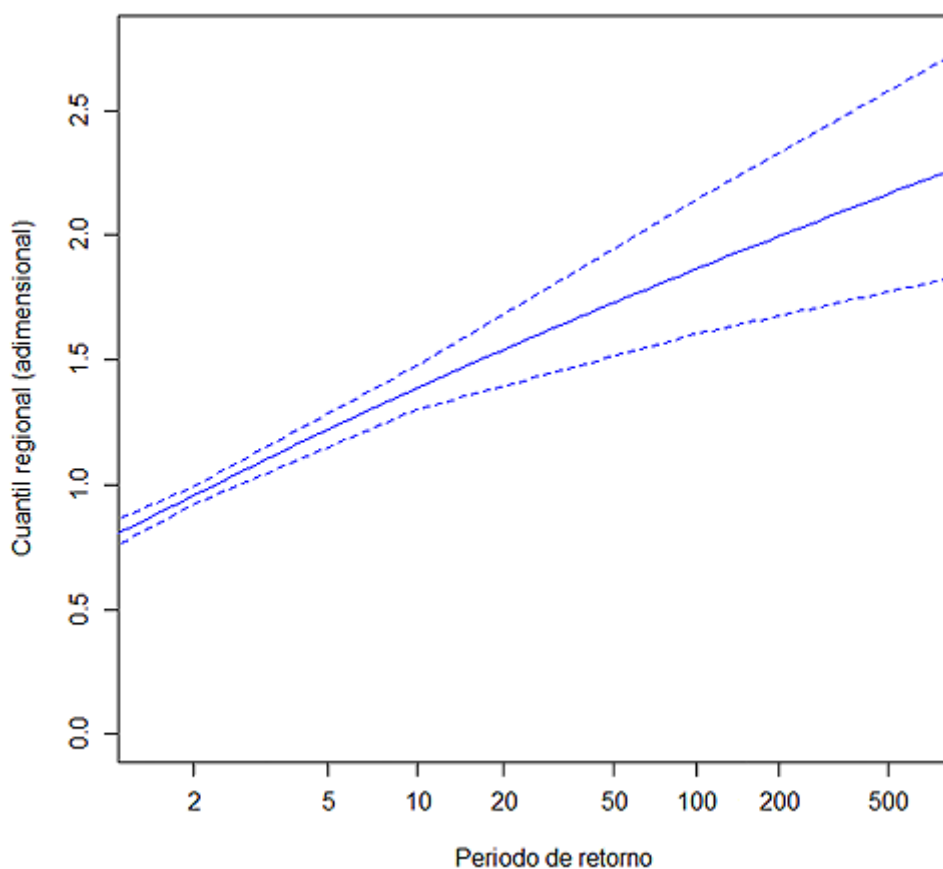


Fig. 3.7. Representación gráfica de las estimaciones regionales de la región Norte.

La tabla 3.16 y la figura 3.7 reflejan cómo la incertidumbre aumenta considerablemente a medida que los cuantiles se hacen más extremos.

Definida la curva regional de frecuencia, se procede a particularizarla localmente en cada estación de la región según la ecuación 2.25:

$$\hat{Q}_i(F) = l_1^{(i)} \hat{q}(F)$$

Los valores regionales adimensionales estimados, $\hat{q}(F)$, se multiplican por el índice de avenida de cada sitio, asumido como la media de sus observaciones, $l_1^{(i)}$.

Tabla 3.17. *Índice de avenida de las estaciones de la región Norte.*

Belzunce	Erro	Espinal	Eugui	Iraizoz	Iroz	Olagüe	Velate	Zubiri
58.55	63.00	62.31	68.03	62.31	46.37	56.79	67.08	56.10

Se presentan a continuación las estimaciones particularizadas en cada estación.

BELZUNCE

Tabla 3.18. Cuantiles de precipitación en 24 h. de Belzunce.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	55.91	4.2348	48.73	61.97
5	71.48	5.4472	61.94	79.80
10	81.24	7.4163	68.07	94.25
20	90.18	8.1197	75.16	104.18
50	101.23	12.1895	79.77	116.85
100	109.18	14.1136	87.68	129.87
200	116.88	17.5206	90.98	145.86
500	126.76	20.9530	93.85	158.61
1000	134.06	22.1148	100.57	170.49

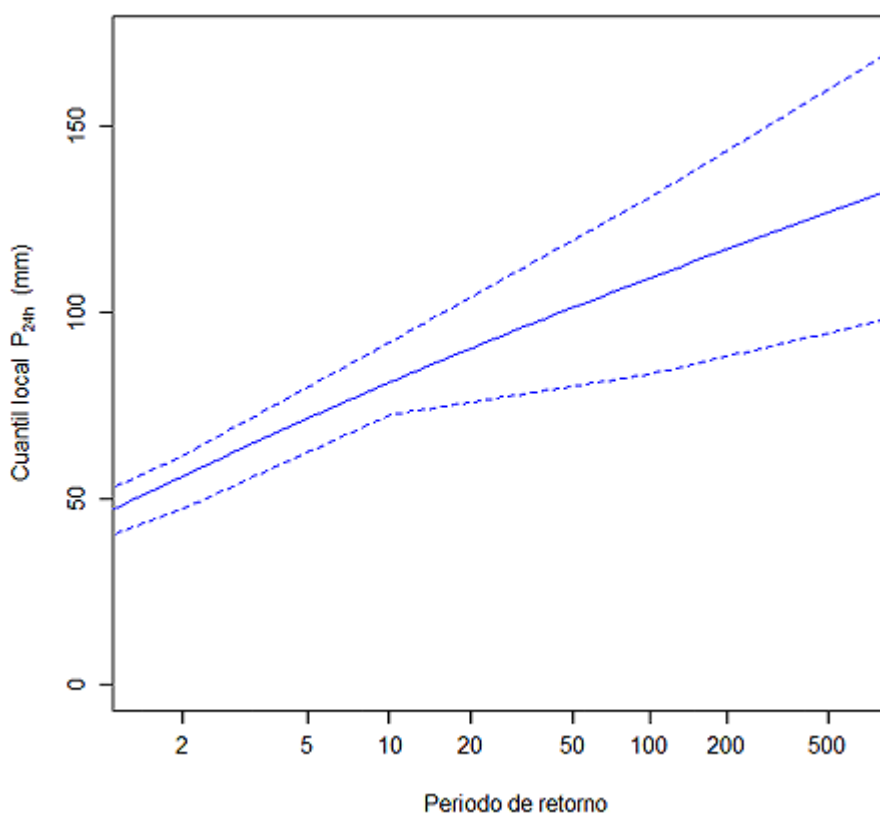


Fig. 3.8. Representación gráfica de los cuantiles de precipitación en 24 h. de Belzunce.

ERRO

Tabla 3.19. Cuantiles de precipitación en 24 h. de Erro.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	60.16	3.2952	55.32	66.69
5	76.92	4.3489	70.30	84.15
10	87.42	4.9997	79.05	95.65
20	97.04	7.3051	85.20	109.37
50	108.93	9.4205	93.98	128.62
100	117.49	10.6676	100.85	136.21
200	125.78	13.4984	102.92	151.24
500	136.41	16.4773	110.95	167.78
1000	144.26	20.0733	115.05	173.34

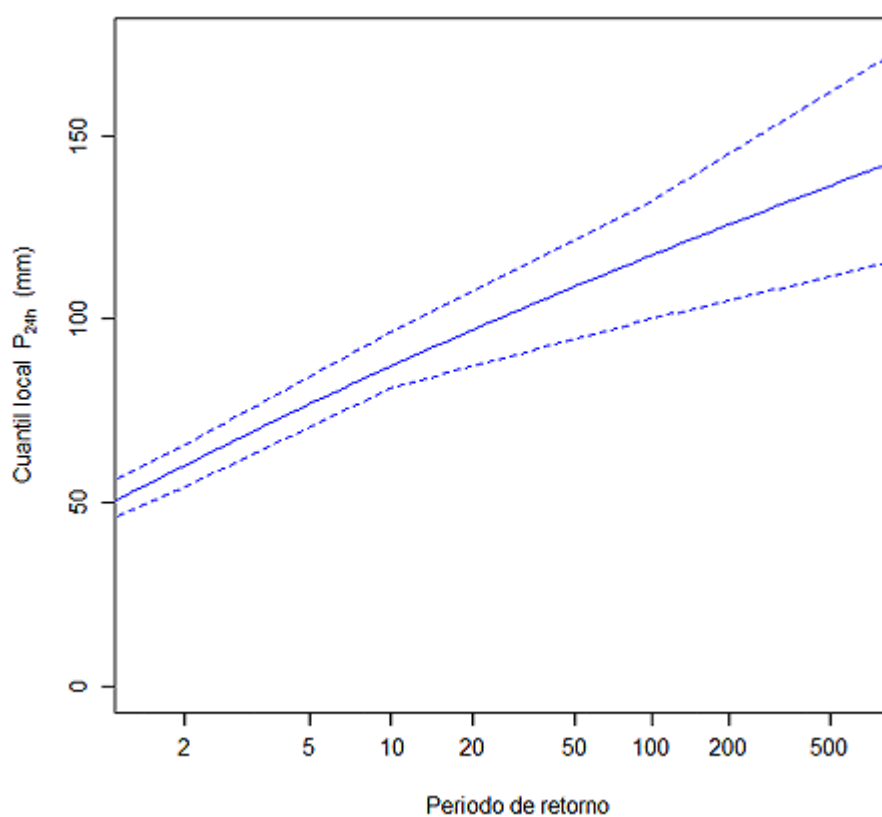


Fig. 3.9. Representación gráfica de los cuantiles de precipitación en 24 h. de Erro.

ESPINAL

Tabla 3.20. Cuantiles de precipitación en 24 h. de Espinal.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	59.50	3.2728	54.45	65.55
5	76.07	4.2276	68.89	83.66
10	86.46	5.3369	78.21	95.51
20	95.98	6.7726	86.16	107.99
50	107.73	9.8274	93.54	122.71
100	116.20	11.6933	96.43	138.60
200	124.40	14.1862	103.35	152.42
500	134.91	15.8330	113.46	167.14
1000	142.68	19.3581	112.14	176.48

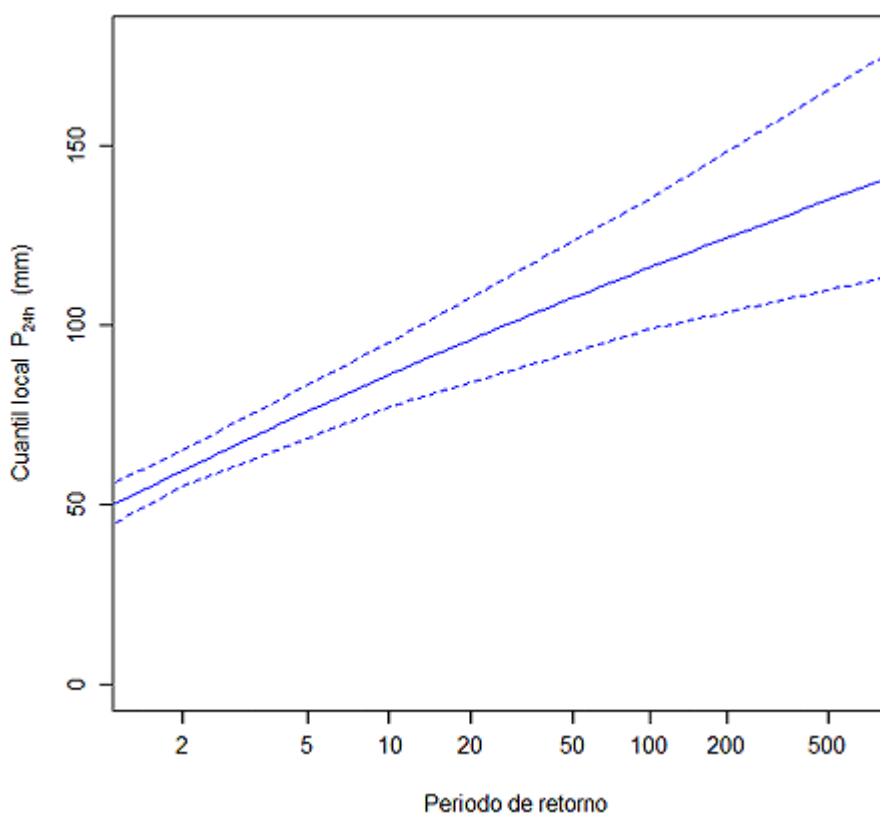


Fig. 3.10. Representación gráfica de los cuantiles de precipitación en 24 h. de Espinal.

EUGUI

Tabla 3.21. Cuantiles de precipitación en 24 h. de Eugui.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	64.97	2.7086	60.39	69.61
5	83.06	4.1583	76.24	89.12
10	94.41	4.9703	87.37	102.98
20	104.80	6.9892	92.96	116.82
50	117.63	10.2891	99.93	135.54
100	126.88	11.9408	105.81	148.77
200	135.82	13.6647	114.00	165.56
500	147.31	17.3460	119.94	181.40
1000	155.78	18.6417	127.83	192.27

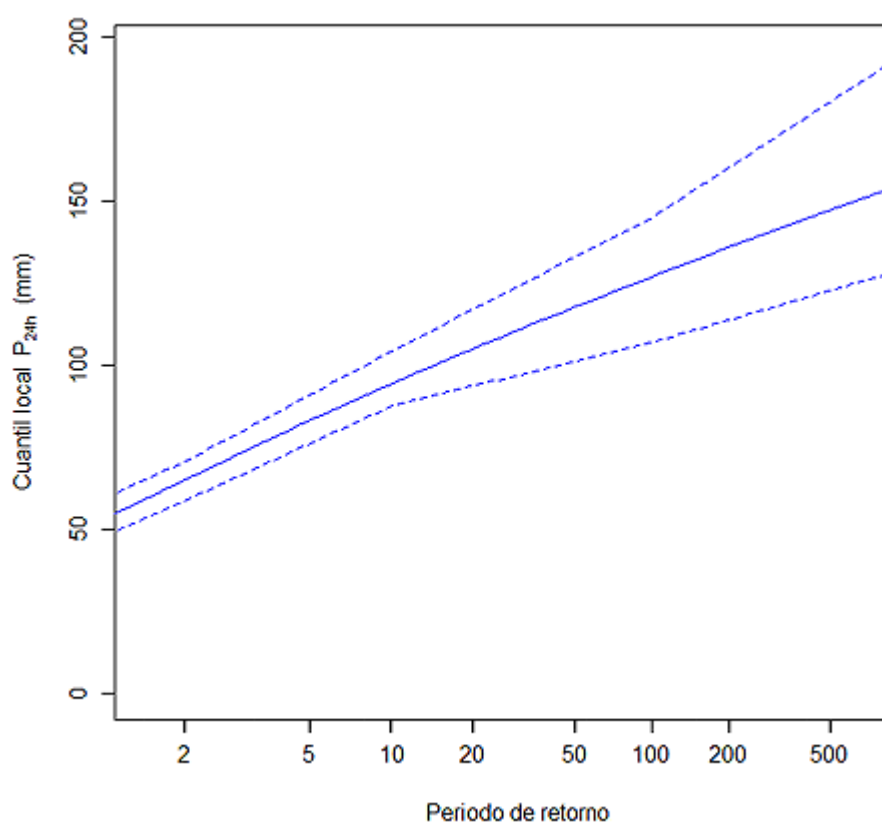


Fig. 3.11. Representación gráfica de los cuantiles de precipitación en 24 h. de Eugui.

IRAIZOZ

Tabla 3.22. Cuantiles de precipitación en 24 h. de Iraizoz.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	59.50	3.4532	53.63	64.75
5	76.08	4.5135	67.66	82.22
10	86.46	5.7167	78.17	97.82
20	95.98	7.5947	83.46	107.46
50	107.74	10.3322	91.90	130.23
100	116.20	11.6899	96.10	135.31
200	124.40	14.6677	99.00	152.34
500	134.92	16.5418	110.20	173.14
1000	142.68	18.9738	116.21	183.08

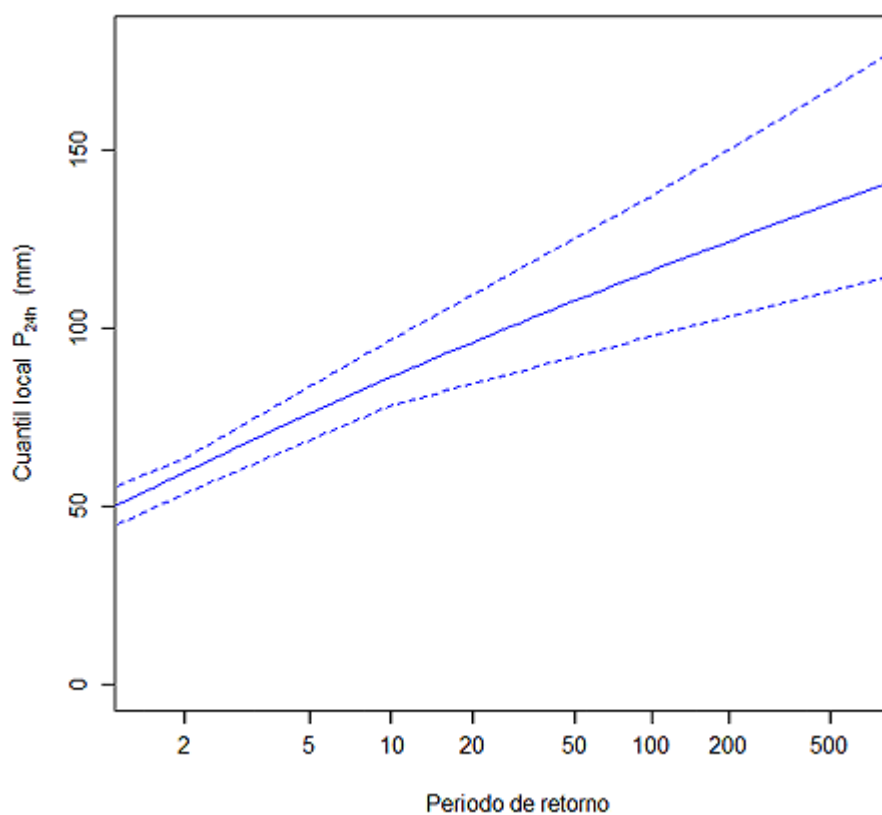


Fig. 3.12. Representación gráfica de los cuantiles de precipitación en 24 h. de Iraizoz.

IROZ

Tabla 3.23. Cuantiles de precipitación en 24 h. de Iroz.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	44.28	2.6207	40.95	49.48
5	56.62	3.5114	51.14	63.18
10	64.35	4.1565	57.38	71.05
20	71.43	5.7783	62.74	81.83
50	80.18	7.8488	68.98	95.72
100	86.48	9.4843	71.54	103.01
200	92.58	11.2336	75.83	112.58
500	100.41	13.8110	78.02	123.96
1000	106.19	16.5181	84.17	132.31

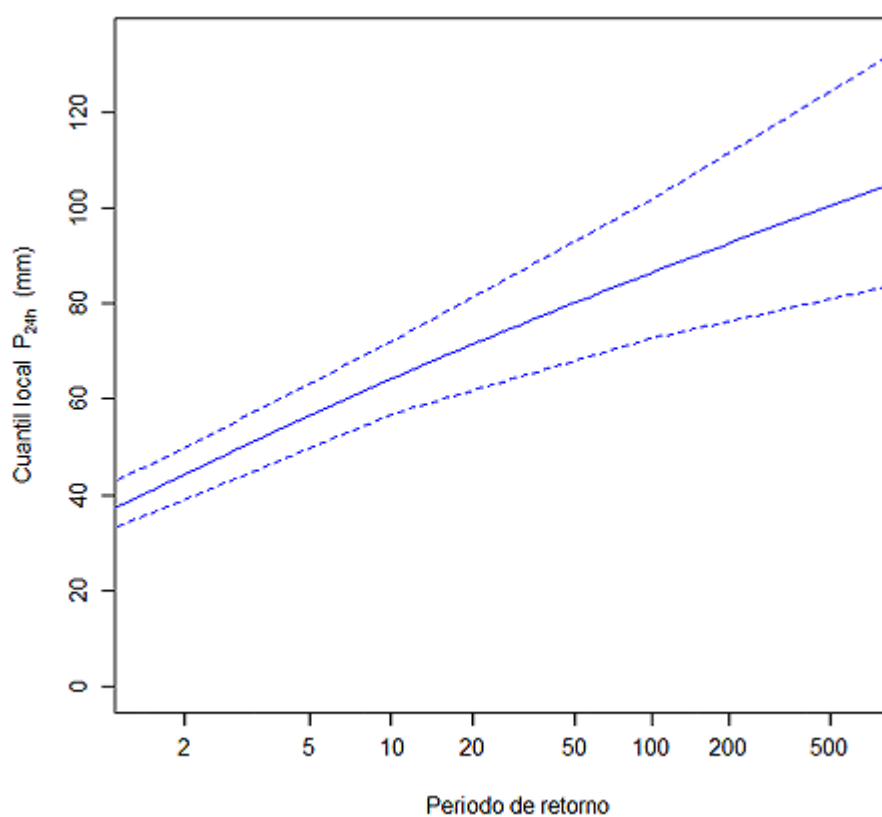


Fig. 3.13. Representación gráfica de los cuantiles de precipitación en 24 h. de Iroz.

OLAGÜE

Tabla 3.24. Cuantiles de precipitación en 24 h. de Olagüe.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	54.23	2.7876	49.86	59.09
5	69.34	4.1528	63.12	76.92
10	78.81	5.5876	69.49	88.22
20	87.48	7.0003	75.89	100.25
50	98.20	9.8500	82.20	116.83
100	105.92	10.4683	89.42	124.71
200	113.39	14.7365	89.73	139.02
500	122.97	16.4186	97.34	152.40
1000	130.05	18.7577	99.44	162.60

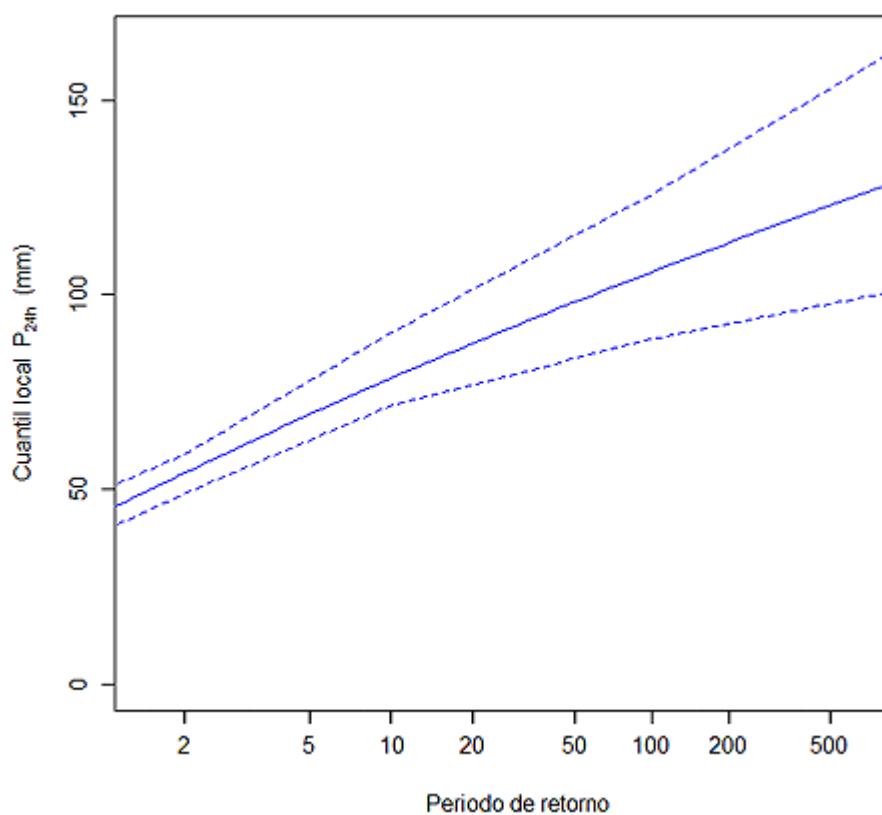


Fig. 3.14. Representación gráfica de los cuantiles de precipitación en 24 h. de Olagüe.

VELATE

Tabla 3.25. Cuantiles de precipitación en 24 h. de Velate.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	64.06	4.7057	57.51	71.78
5	81.90	6.3404	73.58	93.13
10	93.08	7.3703	82.74	106.30
20	103.33	9.3355	89.60	119.94
50	115.99	11.8721	96.01	137.59
100	125.10	14.7167	98.13	152.28
200	133.93	16.3299	110.69	161.16
500	145.25	19.7160	116.07	180.69
1000	153.60	20.5217	121.82	185.63

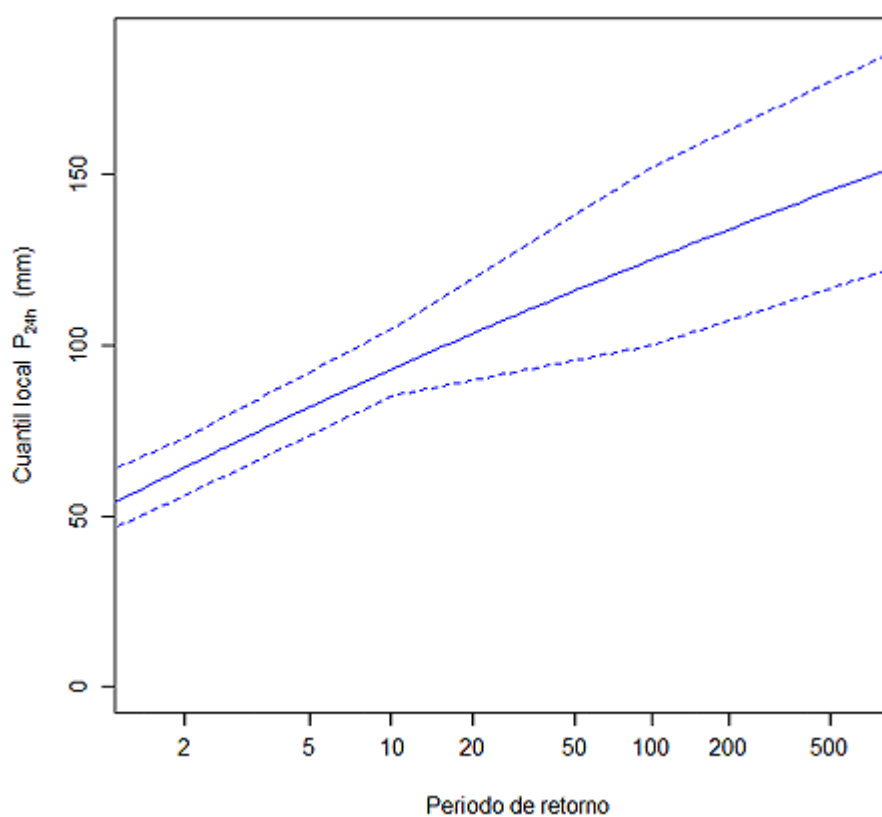


Fig. 3.15. Representación gráfica de los cuantiles de precipitación en 24 h. de Velate.

ZUBIRI

Tabla 3.26. Cuantiles de precipitación en 24 h. de Zubiri.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	53.57	2.5140	48.84	57.93
5	68.49	3.6057	62.87	74.35
10	77.85	5.0332	69.53	86.35
20	86.42	6.2806	76.82	97.12
50	97.00	8.9665	82.25	112.82
100	104.62	9.5126	91.17	124.23
200	112.00	13.6535	89.37	133.37
500	121.47	15.0735	100.05	153.63
1000	128.46	18.0472	102.87	168.03

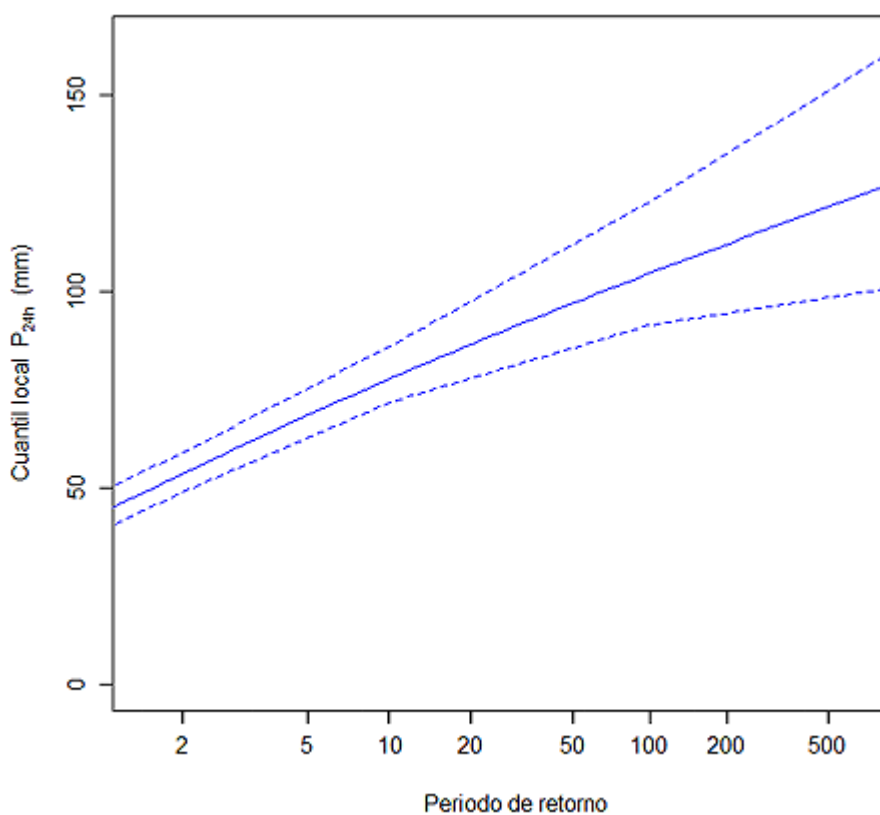


Fig. 3.16. Representación gráfica de los cuantiles de precipitación en 24 h. de Zubiri.

Se reflejan los valores de los cuantiles de precipitación en 24 h. de las diferentes estaciones de la región en el siguiente cuadro resumen:

Tabla 3.27. Cuantiles de precipitación en 24 h. de las estaciones de la región Norte.

T (años)	Cuantil local P _{24h} (mm)								
	Belzunce	Erro	Espinal	Eugui	Iraizoz	Iroz	Olagüe	Velate	Zubiri
2	55.91	60.16	59.50	64.97	59.50	44.28	54.23	64.06	53.57
5	71.48	76.92	76.07	83.06	76.08	56.62	69.34	81.90	68.49
10	81.24	87.42	86.46	94.41	86.46	64.35	78.81	93.08	77.85
20	90.18	97.04	95.98	104.80	95.98	71.43	87.48	103.33	86.42
50	101.23	108.93	107.73	117.63	107.74	80.18	98.20	115.99	97.00
100	109.18	117.49	116.20	126.88	116.20	86.48	105.92	125.10	104.62
200	116.88	125.78	124.40	135.82	124.40	92.58	113.39	133.93	112.00
500	126.76	136.41	134.91	147.31	134.92	100.41	122.97	145.25	121.47
1000	134.06	144.26	142.68	155.78	142.68	106.19	130.05	153.60	128.46

Apréciase cómo, evidentemente, los valores mayores corresponden a las estaciones con mayor índice de avenida, por ser éste el factor de escala. Eugui y Velate constituyen los sitios de mayor precipitación en 24 h. en la región Norte. Por el contrario, Iroz arroja valores notablemente inferiores al resto, dada su mayor cercanía a la cuenca de Pamplona.

3.1.4.1.- REGIÓN SUR

Pearson tipo III (PE3). Parámetros (3): μ (localización), σ (escala), γ (forma).

Siendo $\alpha = 4/\gamma^2$, $\beta = \frac{1}{2}\sigma|\gamma|$, $\xi = \mu - 2\sigma/\gamma$, su función de distribución se expresa como:

$$F(x) = \begin{cases} G\left(\alpha, \frac{x-\xi}{\beta}\right) / \Gamma(\alpha), & \gamma > 0 \\ 1 - G\left(\alpha, \frac{\xi-x}{\beta}\right) / \Gamma(\alpha), & \gamma < 0 \end{cases}$$

siendo $\Gamma(\cdot)$ la función Gamma y $G(\cdot)$ la función Gamma incompleta.

Nuevamente se comienza por estimar los parámetros de la curva regional de frecuencia de la distribución, igualando los ratios medios regionales de L -momentos de las observaciones a los de la distribución:

Tabla 3.28. *Parámetros de la curva regional de frecuencia de la distribución Pearson tipo III de la región Sur.*

μ	σ	γ
1	0.3869	1.4950

Se computan los cuantiles regionales adimensionales de precipitación en 24 h. para los periodos de retorno T de 2, 5, 10, 20, 50, 100, 200, 500 y 1000 años. Se evalúa asimismo la incertidumbre de su estimación mediante el cálculo del error cuadrático medio absoluto y los límites de error al 90%, establecidos en la ecuación 2.34:

$$\frac{\hat{Q}(F)}{U_{.05}(F)} \leq Q(F) \leq \frac{\hat{Q}(F)}{L_{.05}(F)}$$

De nuevo se atiende a la recomendación de Hosking y Wallis (1997), empleándose 1000 simulaciones de una región con número de estaciones y longitudes muestrales similares a las de la observada, y con distribución de frecuencia Pearson tipo III, pero cuyos ratios de L -momentos locales son definidos arbitrariamente en un rango menor que el observado para aproximar a los valores poblacionales:

Tabla 3.29. Resultados de las estimaciones regionales de la región Sur.

T (años)	Cuantil regional	ECM	Límite 0.05	Límite 0.95
2	0.91	0.0218	0.86	0.95
5	1.27	0.0214	1.22	1.31
10	1.52	0.0529	1.41	1.63
20	1.75	0.0889	1.59	1.95
50	2.06	0.1398	1.80	2.37
100	2.29	0.1800	1.96	2.70
200	2.51	0.2212	2.11	3.02
500	2.80	0.2770	2.31	3.46
1000	3.02	0.3200	2.46	3.79

Se representan gráficamente los resultados de las estimaciones, indicándose en trazo discontinuo los límites de error:

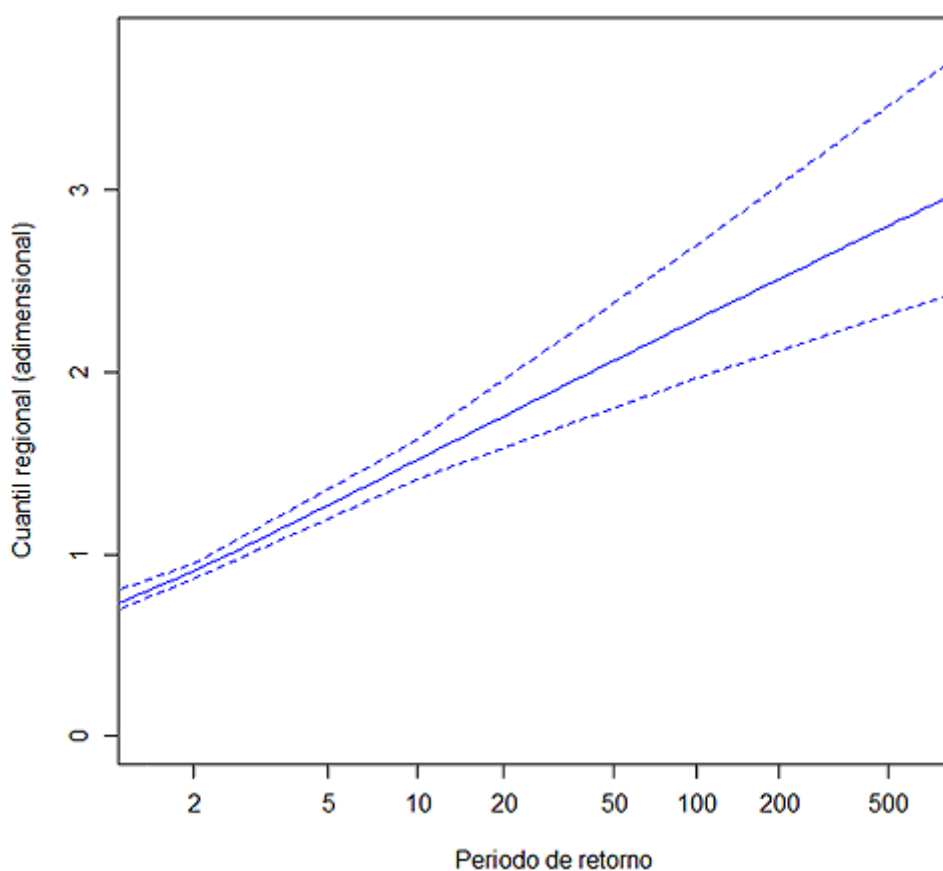


Fig. 3.17. Representación gráfica de las estimaciones regionales de la región Sur.

La tabla 3.29 y la figura 3.17 vuelven a reflejar cómo la incertidumbre es más acusada en los cuantiles más extremos.

Definida la curva regional de frecuencia, se procede a particularizarla localmente en cada estación de la región según la ecuación 2.25:

$$\hat{Q}_i(F) = l_1^{(i)} \hat{q}(F)$$

Los valores regionales adimensionales estimados, $\hat{q}(F)$, se multiplican por el índice de avenida de cada sitio, asumido como la media de sus observaciones, $l_1^{(i)}$.

Tabla 3.30. *Índice de avenida de las estaciones de la región Sur.*

Ilundáin	Monreal	Noáin	Otazu	Pamplona	Zuazu
49.84	49.27	46.19	46.92	52.96	50.40

Se presentan a continuación las estimaciones particularizadas en cada estación.

ILUNDÁIN

Tabla 3.31. Cuantiles de precipitación en 24 h. de Ilundáin.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	45.23	3.2473	39.67	51.10
5	63.17	4.5919	56.99	70.70
10	75.56	6.3364	66.34	89.17
20	87.45	8.2755	76.95	106.58
50	102.71	11.3804	86.72	125.17
100	114.01	13.7728	97.33	145.39
200	125.16	16.1837	101.67	159.97
500	139.72	20.1431	111.56	191.92
1000	150.63	21.7448	125.40	201.57

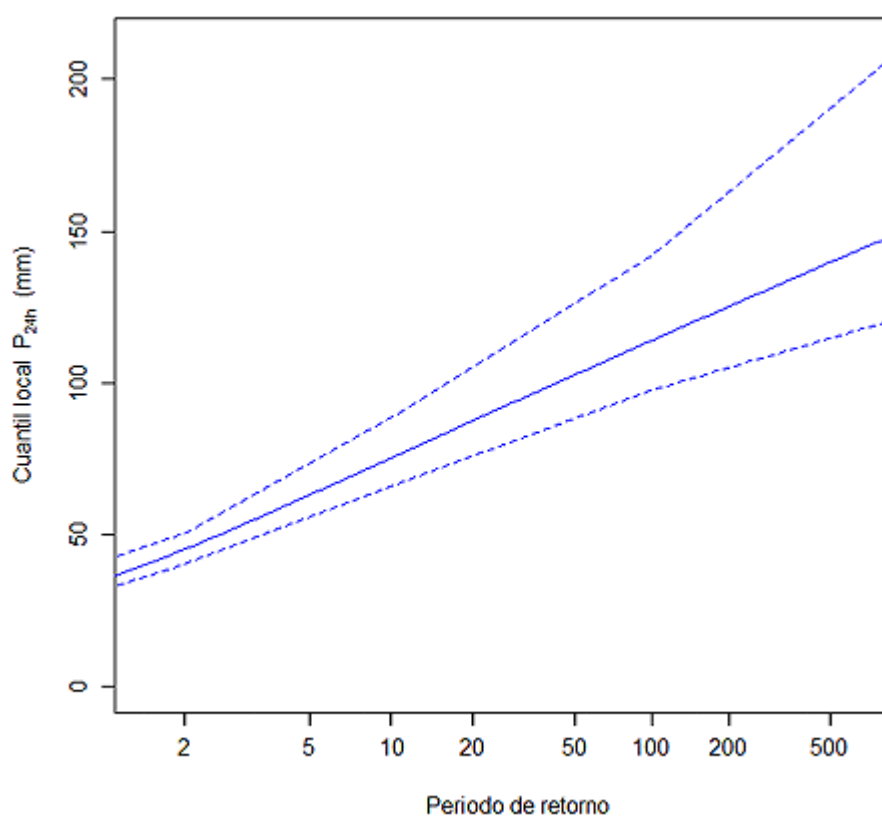


Fig. 3.18. Representación gráfica de los cuantiles de precipitación en 24 h. de Ilundáin.

MONREAL

Tabla 3.32. Cuantiles de precipitación en 24 h. de Monreal.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	44.71	3.3228	39.48	50.22
5	62.44	4.3326	55.72	69.86
10	74.69	5.9993	66.02	85.23
20	86.44	8.1644	75.65	105.23
50	101.53	11.3748	85.94	124.13
100	112.70	12.3740	94.47	140.39
200	123.72	15.4839	104.34	156.02
500	138.12	20.8982	107.26	184.02
1000	148.90	22.5150	114.65	190.96

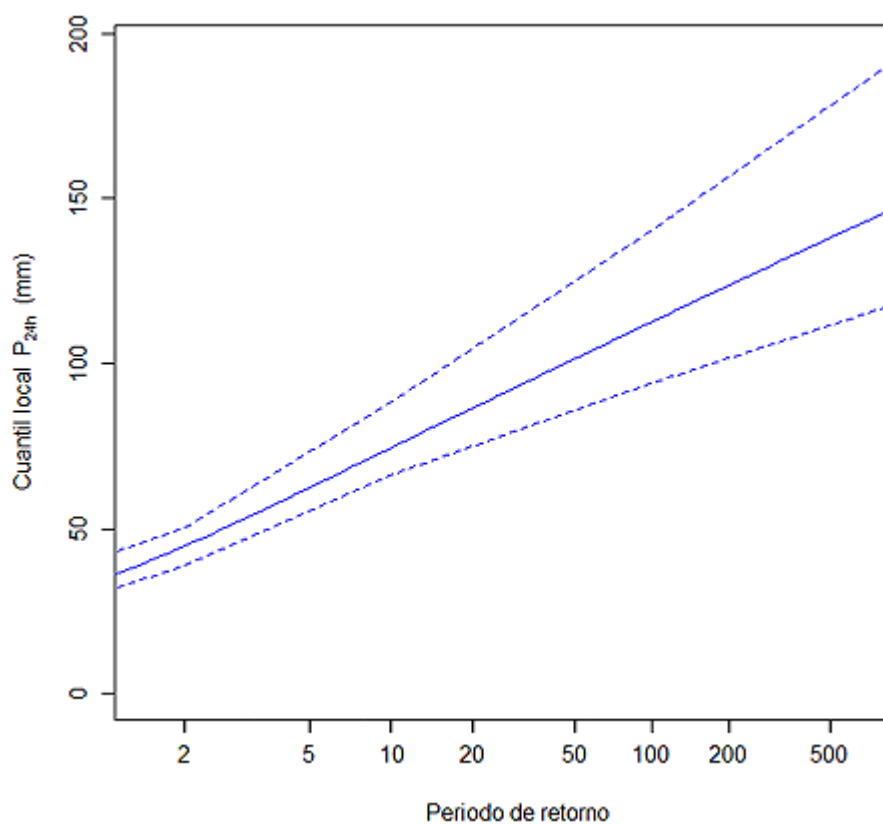


Fig. 3.19. Representación gráfica de los cuantiles de precipitación en 24 h. de Monreal.

NOÁIN

Tabla 3.33. Cuantiles de precipitación en 24 h. de Noáin.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	41.92	3.0054	37.11	47.34
5	58.55	4.0491	52.48	65.07
10	70.03	5.7789	59.61	80.95
20	81.05	7.7990	68.54	95.15
50	95.19	11.4675	76.95	115.52
100	105.67	12.0755	87.51	127.67
200	116.01	14.9039	94.63	141.49
500	129.50	18.9447	102.39	164.34
1000	139.62	22.5998	107.23	177.22

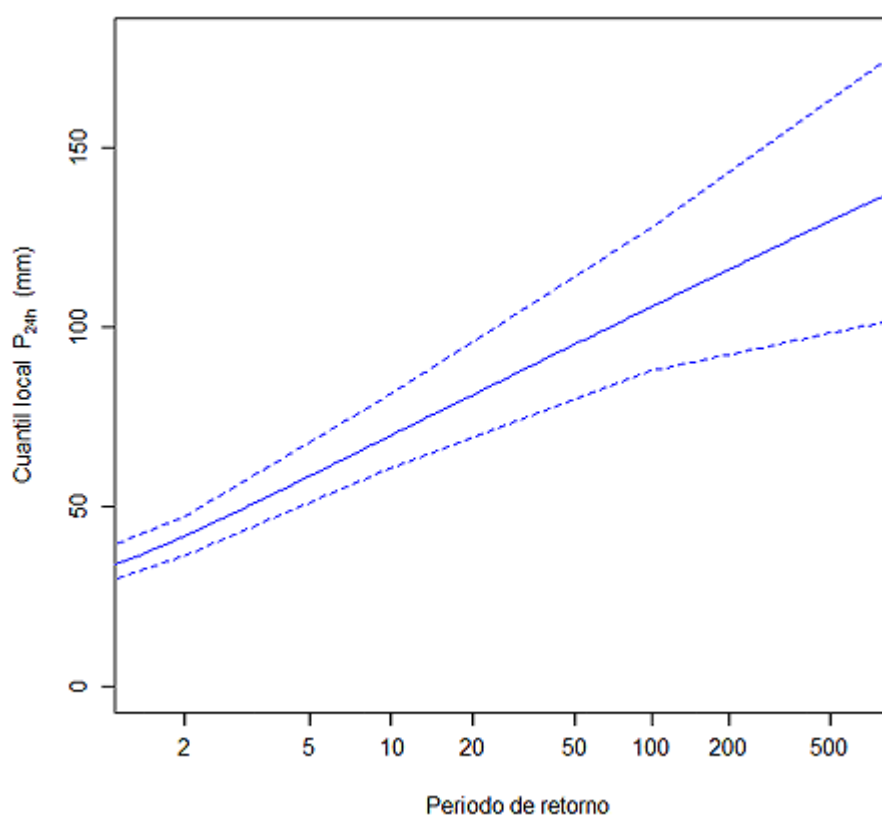


Fig. 3.20. Representación gráfica de los cuantiles de precipitación en 24 h. de Noáin.

OTAZU

Tabla 3.34. Cuantiles de precipitación en 24 h. de Otazu.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	42.57	3.1263	37.17	47.39
5	59.47	4.0863	52.96	66.91
10	71.13	5.8280	62.23	81.49
20	82.32	8.1173	69.15	95.91
50	96.69	10.9411	79.59	119.31
100	107.33	14.1716	86.06	136.34
200	117.82	17.2463	91.67	150.53
500	131.53	20.2954	104.05	181.22
1000	141.80	23.7778	106.79	183.98

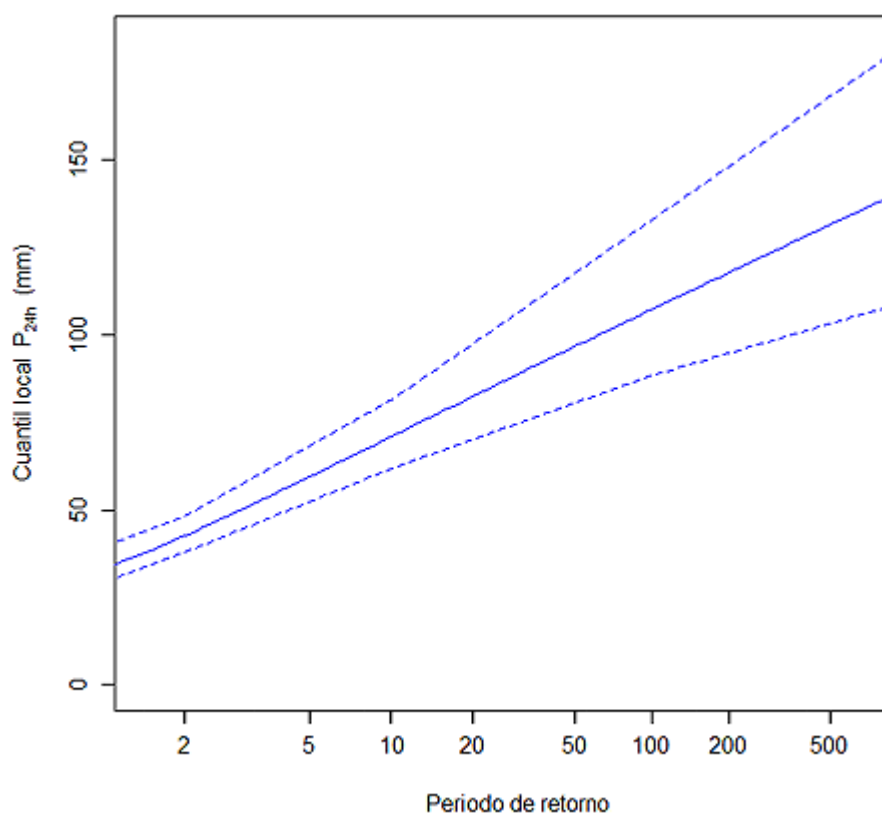


Fig. 3.21. Representación gráfica de los cuantiles de precipitación en 24 h. de Otazu.

PAMPLONA

Tabla 3.35. Cuantiles de precipitación en 24 h. de Pamplona.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	48.06	2.1677	44.38	52.07
5	67.12	2.6680	62.29	71.94
10	80.28	3.8124	73.81	85.78
20	92.92	5.5382	84.05	102.41
50	109.13	8.8110	96.43	122.62
100	121.14	10.7125	103.55	137.98
200	132.99	12.8514	115.26	155.28
500	148.47	14.8753	126.71	171.95
1000	160.06	17.6019	136.62	188.12

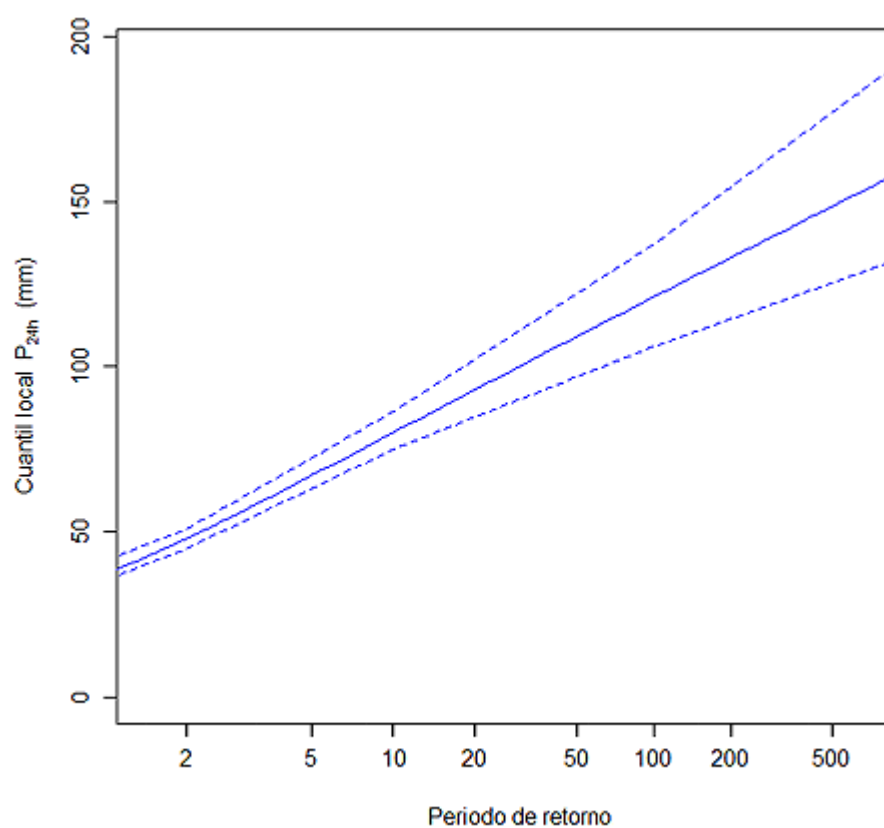


Fig. 3.22. Representación gráfica de los cuantiles de precipitación en 24 h. de Pamplona.

ZUAZU

Tabla 3.36. Cuantiles de precipitación en 24 h. de Zuazu.

T (años)	Cuantil local P_{24h} (mm)	ECM	Límite 0.05	Límite 0.95
2	45.73	3.1937	40.01	50.46
5	63.88	4.5970	55.73	71.22
10	76.41	6.0823	66.80	86.43
20	88.43	8.0786	74.82	102.81
50	103.86	11.2805	86.44	124.45
100	115.29	13.0765	93.99	135.59
200	126.57	14.8203	104.12	153.21
500	141.29	19.5714	113.05	181.86
1000	152.33	21.2519	119.24	192.34

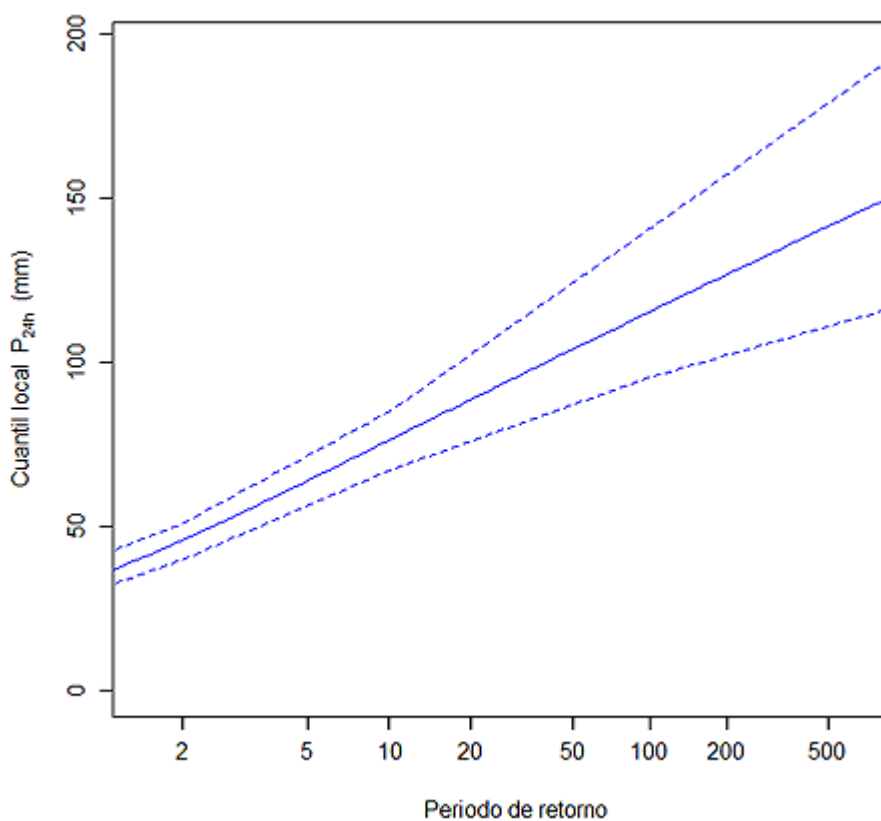


Fig. 3.23. Representación gráfica de los cuantiles de precipitación en 24 h. de Zuazu.

Se reflejan los valores de los cuantiles de precipitación en 24 h. de las diferentes estaciones de la región en el siguiente cuadro resumen:

Tabla 3.37. Cuantiles de precipitación en 24 h. de las estaciones de la región Sur.

T (años)	Cuantil local P_{24h} (mm)					
	Ilundáin	Monreal	Noáin	Otazu	Pamplona	Zuazu
2	45.23	44.71	41.92	42.57	48.06	45.73
5	63.17	62.44	58.55	59.47	67.12	63.88
10	75.56	74.69	70.03	71.13	80.28	76.41
20	87.45	86.44	81.05	82.32	92.92	88.43
50	102.71	101.53	95.19	96.69	109.13	103.86
100	114.01	112.70	105.67	107.33	121.14	115.29
200	125.16	123.72	116.01	117.82	132.99	126.57
500	139.72	138.12	129.50	131.53	148.47	141.29
1000	150.63	148.90	139.62	141.80	160.06	152.33

Pamplona y Zuazu constituyen las estaciones de mayor precipitación en 24 h. en la región Sur, dado que su índice de avenida es superior al del resto. Por el contrario, los menores valores recaen en este caso en Noáin.

3.2.- COMPARACIÓN CON LOS MÉTODOS DE ANÁLISIS LOCAL

Estimados los cuantiles de precipitación en 24 h. mediante el estudio regional, se procede a su comparación con los métodos clásicos de análisis local de cara a valorar la robustez del modelo.

Se opta por ajustar un total de cuatro distribuciones a cada estación mediante el método tradicional de los momentos. Por un lado, se contrasta la función Pearson tipo III seleccionada en el análisis regional. Además, se evalúan tres de las distribuciones más habituales en el estudio de eventos extremos locales según el Centro de Estudios Hidrográficos del CEDEX: Valores Extremos tipo I o Gumbel, de frecuente aplicación en España; General de Valores Extremos, empleada actualmente en el Reino Unido; y LogPearson III, utilizada en la mayor parte de los trabajos de Estados Unidos.

Se presentan las funciones de distribución de Gumbel y LogPearson III:

- Gumbel. Parámetros (2): μ (localización), σ (escala).

$$F(x) = e^{-e^{-y}} \quad \text{siendo } y = \frac{(x - \mu)}{\sigma}$$

- LogPearson III. Parámetros (3): α, β, γ .

$$F(x) = \frac{\Gamma_{(\ln(x)-\gamma)/\beta}(\alpha)}{\Gamma(\alpha)}$$

El procedimiento se ejecuta mediante Hydrognomon (ITIA Research Team, 2010), aplicación de software libre para el análisis y procesamiento de datos hidrológicos, principalmente en forma de series temporales. Se configura para estimar los mismos cuantiles computados en el análisis regional, y definir asimismo los límites de error al 90% mediante simulación tipo Monte Carlo.

Se presentan a continuación los resultados del análisis local de cada estación en cada región.

3.2.1.- REGIÓN NORTE

BELZUNCE

Tabla 3.38. Cuantiles de precipitación en 24 h. de Belzunce por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	53.76	44.67	62.66	55.04	46.18	62.46
5	73.58	60.80	90.11	73.88	61.85	87.77
10	87.02	70.39	111.81	86.35	71.42	105.71
20	99.82	77.97	135.76	98.31	80.05	122.57
50	116.14	86.85	164.33	113.79	90.45	143.44
100	128.18	92.69	189.23	125.39	98.23	159.42
200	140.01	97.79	214.02	136.95	106.02	175.32
500	155.43	103.91	246.16	152.20	116.21	196.20
1000	166.95	-	-	163.73	-	-

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	54.64	46.66	63.12	54.17	47.04	61.57
5	73.14	61.99	88.57	72.89	61.35	85.89
10	85.85	69.80	107.51	86.22	70.09	108.34
20	98.39	76.12	127.01	99.73	78.93	136.21
50	115.16	82.85	154.49	118.34	90.59	185.48
100	128.13	87.04	176.81	133.24	98.07	231.29
200	141.42	91.25	201.29	148.97	105.54	336.07
500	159.51	96.00	239.48	171.25	-	-
1000	173.61	-	-	189.33	-	-

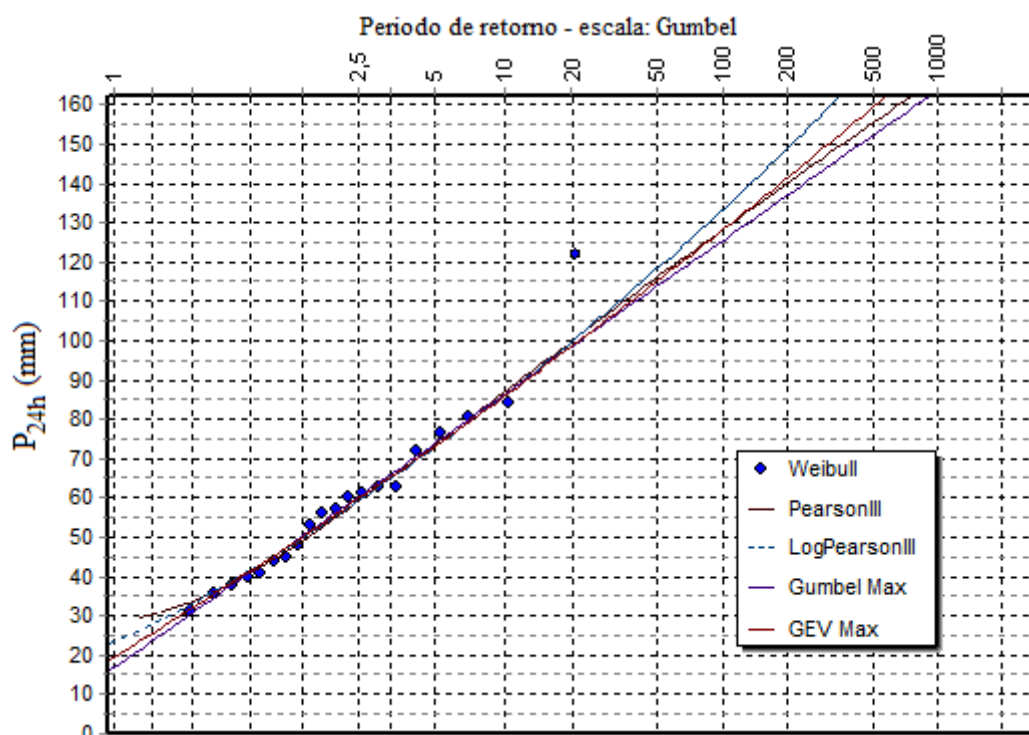


Fig. 3.24. Representación de los cuantiles de precipitación en 24 h. de Belzunce por el método local.

ERRO

Tabla 3.39. Cuantiles de precipitación en 24 h. de Erro por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	60.82	55.25	66.35	60.16	54.88	64.76
5	76.53	69.42	84.12	75.43	67.93	83.70
10	86.05	77.18	96.87	85.53	75.92	96.90
20	94.62	83.69	110.34	95.23	83.24	109.42
50	105.06	91.49	127.09	107.77	92.72	125.57
100	112.48	97.07	138.22	117.18	99.65	137.65
200	119.62	102.03	150.78	126.54	106.42	149.62
500	128.70	108.15	166.28	138.90	115.25	165.50
1000	135.36	112.49	177.52	148.25	125.67	189.41

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	60.82	55.58	66.12	60.82	55.02	65.03
5	76.36	69.45	83.95	76.27	66.60	83.51
10	85.97	76.99	96.17	85.84	74.99	96.69
20	94.72	83.28	109.37	94.65	83.79	113.28
50	105.39	90.70	128.14	105.66	91.86	133.19
100	112.93	95.84	143.39	113.69	99.07	145.62
200	120.07	100.61	157.65	121.58	110.35	154.80
500	128.97	105.86	179.30	131.88	118.69	183.02
1000	135.31	109.13	197.31	139.62	123.95	201.03

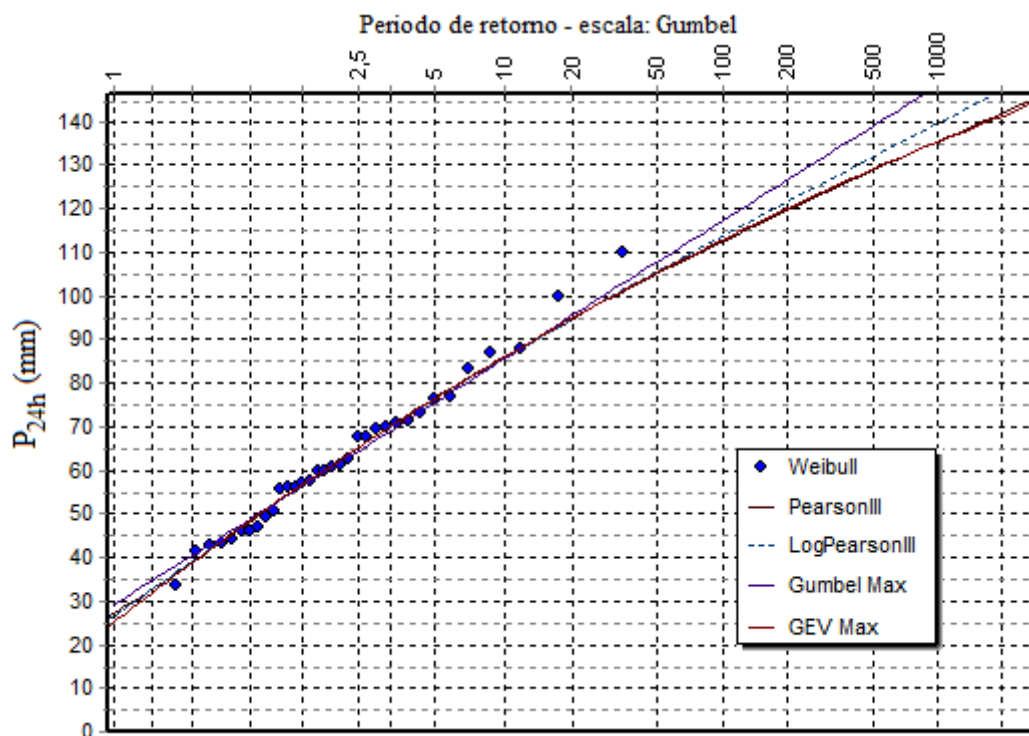


Fig. 3.25. Representación de los cuantiles de precipitación en 24 h. de Erro por el método local.

ESPINAL

Tabla 3.40. Cuantiles de precipitación en 24 h. de Espinal por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	59.98	53.37	66.33	58.77	52.45	64.29
5	79.40	71.05	88.15	77.79	68.87	87.51
10	90.93	80.77	103.38	90.38	78.95	103.78
20	101.21	88.62	118.78	102.45	88.29	119.31
50	113.59	97.84	140.34	118.08	100.24	139.01
100	122.35	104.43	157.89	129.80	108.83	154.02
200	130.71	110.03	172.12	141.47	117.60	168.91
500	141.30	117.19	196.09	156.86	129.03	188.58
1000	149.03	122.28	215.67	168.50	140.13	207.21

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	59.89	53.55	66.26	58.78	52.27	64.20
5	79.29	71.12	88.07	78.90	69.61	88.26
10	91.02	80.86	102.46	92.02	80.96	106.95
20	101.50	88.79	117.60	104.48	91.50	127.32
50	114.03	97.76	137.84	120.54	103.42	157.68
100	122.71	103.70	153.15	132.59	111.35	183.09
200	130.79	109.11	169.42	144.67	122.70	214.72
500	140.66	115.43	190.38	160.81	133.01	254.86
1000	147.55	119.57	205.98	173.18	-	-

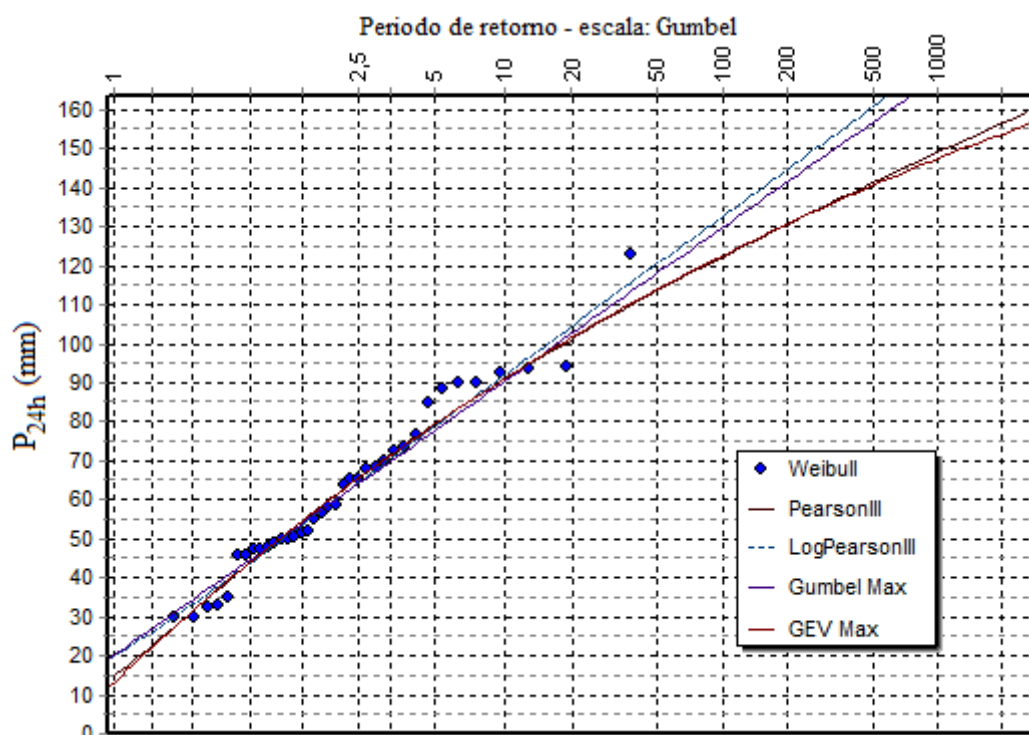


Fig. 3.26. Representación de los cuantiles de precipitación en 24 h. de Espinal por el método local.

EUGUI

Tabla 3.41. Cuantiles de precipitación en 24 h. de Eugui por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	65.29	59.77	70.65	64.98	59.74	69.55
5	82.34	75.00	90.08	81.42	74.06	89.74
10	92.93	83.87	103.81	92.30	82.91	103.74
20	102.58	91.39	117.69	102.74	91.16	117.05
50	114.45	99.79	135.62	116.26	101.55	134.12
100	122.97	105.30	148.43	126.39	109.14	146.84
200	131.20	110.27	161.18	136.48	116.90	159.48
500	141.73	115.96	178.09	149.79	126.89	176.32
1000	149.50	120.65	187.51	159.85	137.54	197.78

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	65.41	60.22	70.54	65.08	60.13	69.79
5	82.07	74.99	89.97	81.72	74.84	89.26
10	92.64	83.44	103.39	92.58	83.76	103.75
20	102.45	90.83	117.03	102.94	91.78	119.88
50	114.69	99.01	136.59	116.39	101.82	144.20
100	123.53	104.67	152.58	126.56	109.11	165.79
200	132.07	109.80	169.90	136.84	115.44	195.00
500	142.94	115.70	194.89	150.70	126.37	230.32
1000	150.86	119.87	214.17	166.91	-	-

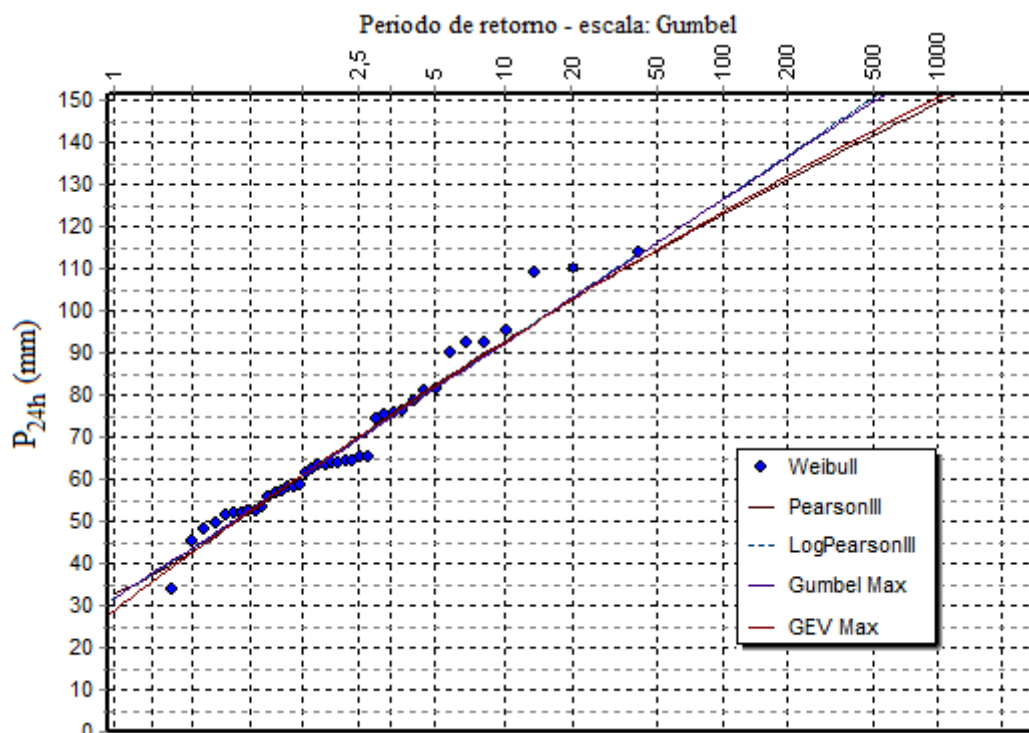


Fig. 3.27. Representación de los cuantiles de precipitación en 24 h. de Eugui por el método local.

IRAIZOZ

Tabla 3.42. Cuantiles de precipitación en 24 h. de Iraizoz por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	59.92	54.75	64.84	59.86	55.00	64.01
5	73.65	66.55	81.07	73.04	66.26	80.43
10	82.29	73.61	92.71	81.77	73.16	91.94
20	90.23	79.22	105.60	90.14	79.49	102.95
50	100.05	85.55	121.76	100.98	87.46	117.03
100	107.14	89.73	134.28	109.10	93.38	127.59
200	114.01	93.60	147.28	117.19	99.28	138.07
500	122.82	98.47	164.76	127.87	107.03	151.87
1000	129.34	102.70	174.59	135.94	115.65	163.21

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	60.08	55.32	64.86	59.32	55.21	63.87
5	73.38	66.81	80.73	72.63	66.12	79.78
10	81.95	73.13	91.90	81.88	72.82	93.43
20	90.00	78.50	104.01	91.10	79.22	109.58
50	100.17	84.70	123.23	103.59	87.40	134.36
100	107.61	88.96	140.98	113.42	93.40	158.22
200	114.88	93.05	159.25	123.68	99.18	188.26
500	124.25	98.10	182.90	138.00	-	-
1000	131.16	100.88	202.10	149.49	-	-

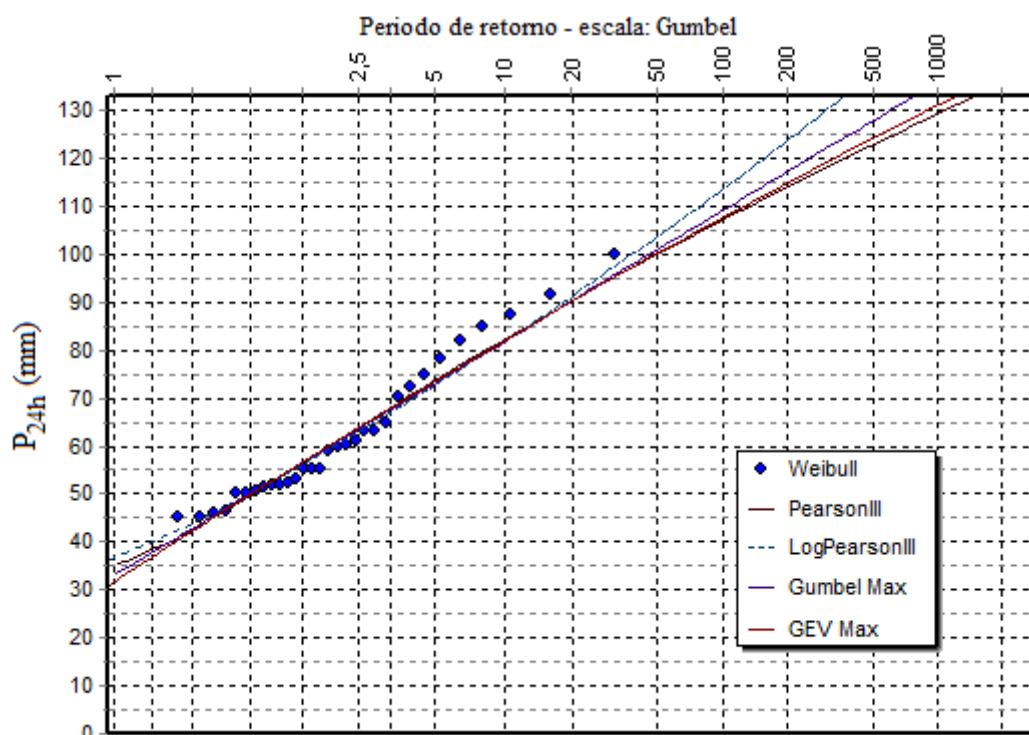


Fig. 3.28. Representación de los cuantiles de precipitación en 24 h. de Iraizoz por el método local.

IROZ

Tabla 3.43. Cuantiles de precipitación en 24 h. de Iroz por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	45.16	39.98	49.80	44.05	39.06	48.29
5	57.80	51.41	64.64	56.57	49.65	64.30
10	65.13	57.65	74.90	64.86	56.11	75.52
20	71.57	62.65	86.15	72.82	61.99	86.21
50	79.23	68.26	107.09	83.11	69.39	99.92
100	84.60	71.97	128.73	90.83	74.86	110.19
200	89.68	75.67	134.99	98.52	80.03	120.45
500	96.08	79.96	143.73	108.66	87.16	133.95
1000	100.72	82.82	153.91	116.33	93.11	145.16

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	45.04	40.08	50.04	43.42	38.86	48.16
5	57.84	51.63	64.56	57.36	50.14	66.05
10	65.33	57.75	74.37	67.10	57.23	81.60
20	71.87	62.64	85.50	76.84	63.88	100.14
50	79.48	68.00	101.60	90.10	73.15	132.69
100	84.61	71.39	114.73	100.58	81.03	162.23
200	89.28	73.89	130.14	111.55	89.12	195.57
500	94.83	77.08	146.17	126.91	-	-
1000	98.61	79.32	156.62	139.26	-	-

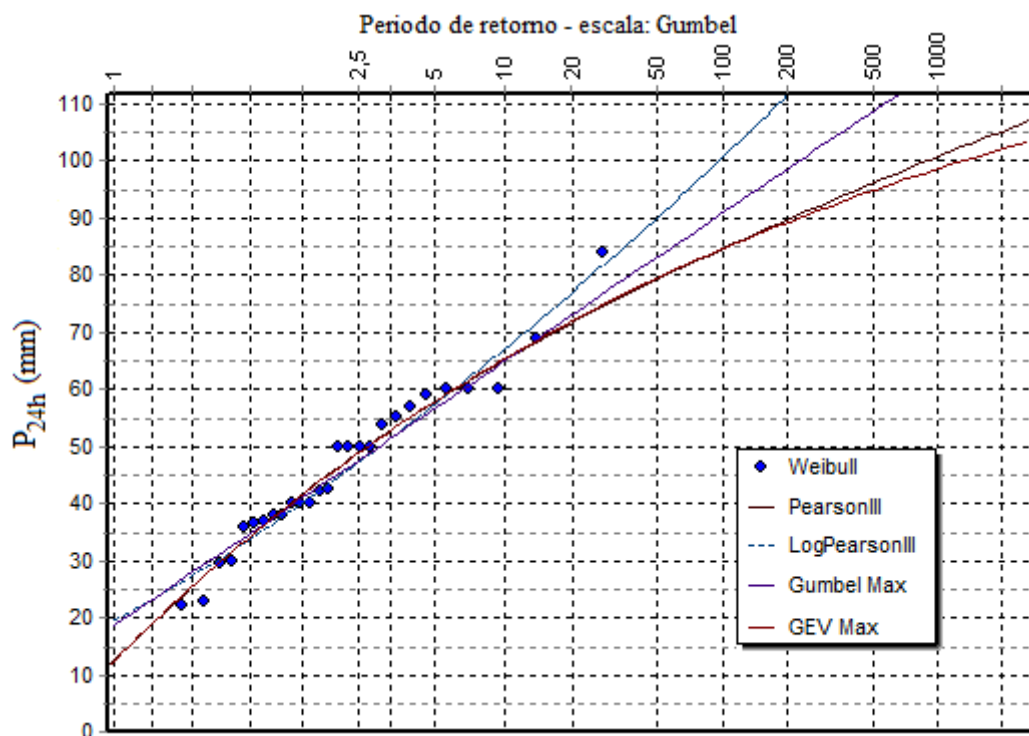


Fig. 3.29. Representación de los cuantiles de precipitación en 24 h. de Iroz por el método local.

OLAGÜE

Tabla 3.44. Cuantiles de precipitación en 24 h. de Olagüe por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	55.03	49.91	60.23	54.17	49.34	58.24
5	69.43	62.71	76.70	68.27	61.41	76.49
10	78.02	69.78	89.49	77.59	68.75	88.81
20	85.68	75.67	100.77	86.54	75.71	100.78
50	94.93	82.65	117.15	98.12	83.93	115.00
100	101.47	87.30	128.07	106.80	90.65	126.93
200	107.73	91.64	140.15	115.45	96.81	137.85
500	115.66	96.91	154.99	126.86	100.41	143.98
1000	121.46	101.54	169.12	135.48	108.32	151.35

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	54.97	50.14	60.48	54.73	49.31	58.93
5	69.34	62.90	76.59	69.02	61.17	76.30
10	78.06	69.83	88.57	77.91	70.19	89.17
20	85.88	75.57	99.10	86.11	77.27	102.31
50	95.24	82.11	121.03	96.38	84.80	120.03
100	101.75	86.47	135.76	103.90	90.87	135.76
200	107.83	89.85	151.60	111.29	97.16	152.81
500	115.28	93.76	171.52	120.96	103.21	179.90
1000	120.50	96.50	185.49	128.24	-	-

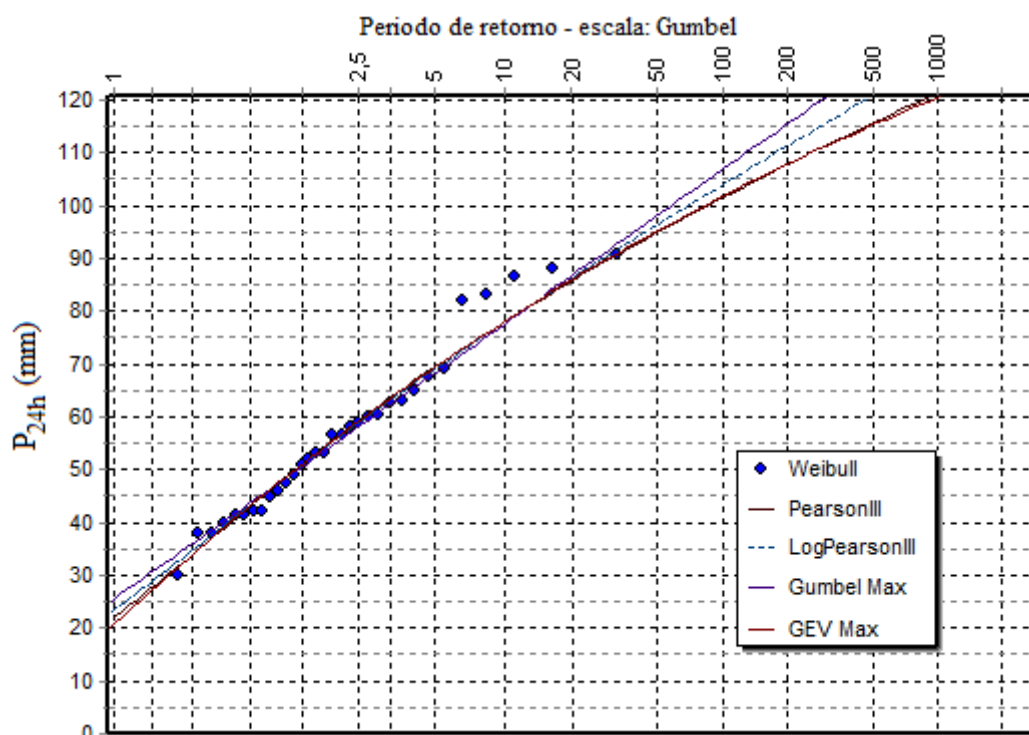


Fig. 3.30. Representación de los cuantiles de precipitación en 24 h. de Olagüe por el método local.

VELATE

Tabla 3.45. Cuantiles de precipitación en 24 h. de Velate por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	67.08	60.05	71.78	64.62	58.51	69.57
5	79.69	73.35	86.08	77.87	69.53	87.25
10	86.29	80.34	93.86	86.64	76.14	99.54
20	91.73	85.33	102.59	95.06	82.16	111.35
50	97.86	90.71	112.90	105.95	89.70	126.21
100	101.95	94.38	118.76	114.11	95.33	137.20
200	105.69	96.40	124.36	122.25	100.91	148.16
500	110.22	99.04	131.92	132.98	105.64	159.01
1000	113.40	101.35	137.84	141.09	108.12	167.39

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	66.62	59.74	73.63	65.47	59.08	70.41
5	79.93	72.19	87.39	79.29	72.47	86.24
10	86.97	78.75	95.83	87.64	79.78	98.08
20	92.64	83.21	106.23	95.20	85.98	110.87
50	98.66	87.56	124.30	104.48	93.70	128.19
100	102.39	90.09	142.02	111.17	98.96	141.07
200	105.52	92.80	159.89	117.67	102.25	154.85
500	108.94	95.66	174.39	126.05	110.49	172.26
1000	111.08	97.20	178.27	132.29	-	-

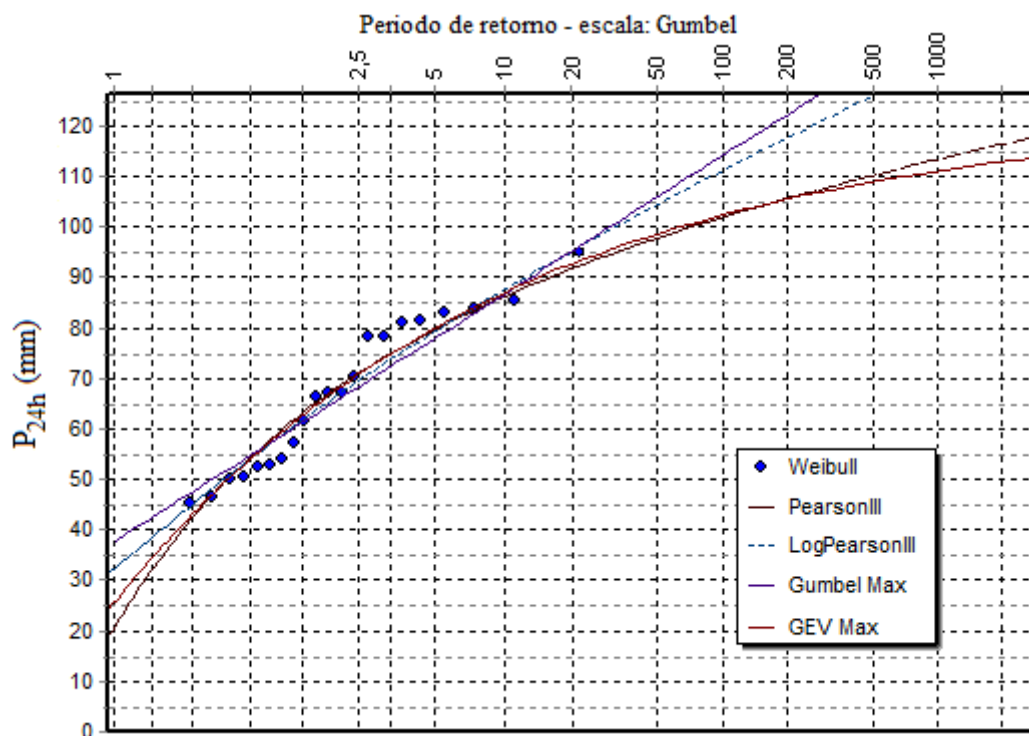


Fig. 3.31. Representación de los cuantiles de precipitación en 24 h. de Velate por el método local.

ZUBIRI

Tabla 3.46. Cuantiles de precipitación en 24 h. de Zubiri por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	54.19	49.59	58.70	53.72	49.46	57.38
5	67.42	61.65	73.60	66.54	60.32	73.18
10	75.48	68.18	84.19	75.03	67.14	84.20
20	82.77	73.69	94.88	83.18	73.41	94.74
50	91.66	80.06	109.52	93.72	81.24	108.16
100	98.01	84.36	121.18	101.62	87.01	118.15
200	104.11	88.37	131.39	109.49	92.74	128.09
500	111.90	93.42	144.46	119.88	100.28	141.45
1000	117.61	97.06	154.18	127.72	106.06	151.45

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	54.21	49.96	58.45	54.36	49.37	57.79
5	67.25	61.56	73.47	67.32	59.70	72.75
10	75.37	67.93	83.66	75.29	68.93	83.93
20	82.81	73.31	94.79	82.57	76.52	93.50
50	91.93	79.52	110.70	91.61	83.51	109.31
100	98.42	83.68	123.65	98.18	86.03	125.91
200	104.60	87.87	134.69	104.61	89.43	135.99
500	112.35	92.21	152.13	112.96	93.32	152.79
1000	117.90	94.76	167.23	119.21	96.51	165.18

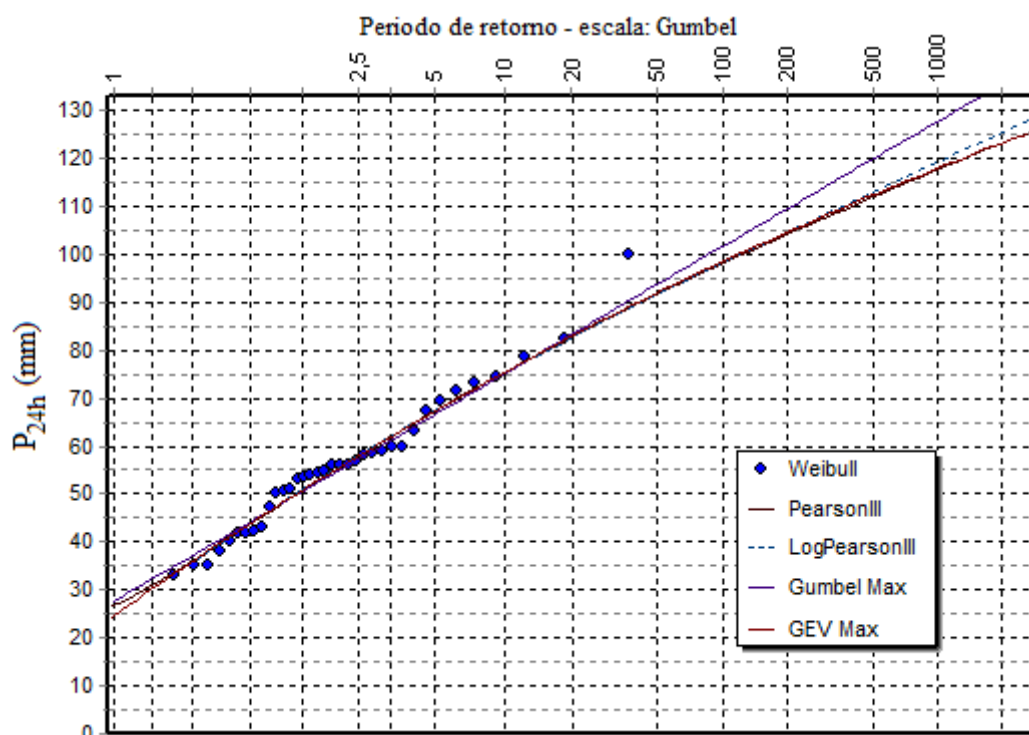


Fig. 3.32. Representación de los cuantiles de precipitación en 24 h. de Zubiri por el método local.

Se analizan los resultados obtenidos para cada estación de la región Norte:

- Belzunce. Obsérvese cómo, salvo para el periodo de retorno de 2 años, los valores de los cuantiles obtenidos regionalmente son considerablemente inferiores a los estimados por el método local. En los periodos de retorno mayores (500 y 1000 años), las estimaciones locales de las distribuciones General de Valores Extremos y LogPearson III se disparan. Por otro lado, la incertidumbre es mayor en el análisis local que en el regional, pues su intervalo definido por los límites de error al 90% resulta mucho más amplio. Dicha incertidumbre se agudiza enormemente en los cuantiles más extremos.
- Erro. En este caso, los valores de los cuantiles obtenidos regionalmente son ligeramente superiores pero muy parecidos a los estimados por el método local. Se comprueba cómo la estimación regional de la distribución Pearson III resulta prácticamente similar a la local de Gumbel. No obstante, la incertidumbre vuelve a ser mayor en el análisis local que en el regional por resultar su intervalo de confianza más amplio, aunque las diferencias no son tan acusadas.
- Espinal. Al igual que en Belzunce, los valores de los cuantiles obtenidos regionalmente son considerablemente inferiores a los estimados por el método local. Las estimaciones locales de las distribuciones LogPearson III y Gumbel se disparan frente al resto, sobretodo en los periodos de retorno mayores (500 y 1000 años). Nuevamente la incertidumbre vuelve a ser mayor en el análisis local que en el regional por resultar su intervalo de confianza mucho más amplio.
- Eugui. En este caso, los valores de los cuantiles obtenidos regionalmente son muy parecidos a los estimados por el método local, resultando ligeramente superiores salvo en los periodos de retorno mayores (500 y 1000 años). La estimación regional de la distribución Pearson III guarda mucha similitud con la aplicación local de la función Gumbel. Una vez más, la incertidumbre vuelve a ser mayor en el análisis local que en el regional por resultar su intervalo más amplio, salvo en los cuantiles más extremos (200, 500 y 1000 años) donde la confianza que ofrece la distribución local Gumbel es superior.
- Iraizoz. Se aprecia cómo, salvo para el periodo de retorno de 1000 años de la distribución local LogPearson III, los valores de los cuantiles obtenidos regionalmente son superiores a los estimados por el método local. En esta ocasión, la incertidumbre del análisis local y el regional resultan parecidas, mejorando el primero la fiabilidad de los cuantiles asociados a los periodos de retorno menores aunque con superioridad del regional en los valores más extremos (500 y 1000 años). Se destaca, no obstante, que la función local Gumbel presenta la menor incertidumbre.
- Iroz. Para los periodos de retorno de 2, 5 y 10 años, las estimaciones resultan bastante similares. A partir del periodo de retorno de 20 años, los cuantiles de la distribución Pearson III del análisis regional superan a los de las distribuciones locales Pearson III y General de Valores Extremos, pero son inferiores a los de Gumbel y LogPearson III. Por otro lado, la incertidumbre resulta mayor en el

análisis local que en el regional, pues sus intervalos de confianza vuelven a ser más amplios.

- Olagüe. Se observa cómo las estimaciones resultan similares hasta el periodo de retorno de 10 años. A partir del periodo de retorno de 20 años, los cuantiles del análisis regional superan a los estimados localmente, salvo en el caso de los más extremos (500 y 1000 años) de la función Gumbel. La incertidumbre resulta mayor en el análisis local que en el regional, con diferencias poco acentuadas.
- Velate. Salvo para el periodo de retorno de 2 años, los valores de los cuantiles obtenidos regionalmente superan notablemente a los estimados por el método local. Sin embargo, en esta ocasión la incertidumbre del análisis local es menor que la del regional, y las diferencias más acusadas las refleja precisamente la distribución Pearson tipo III aplicada en ambas metodologías.
- Zubiri. En este caso, los cuantiles obtenidos regionalmente resultan superiores a los estimados localmente. En lo que respecta a la incertidumbre, el análisis regional y el local arrojan resultados parecidos para los periodos de retorno de 2, 5, 10 y 20 años. Sin embargo, a partir del periodo de retorno de 200 años, la incertidumbre de la distribución Pearson tipo III aplicada regionalmente, si bien es inferior a la de la General de Valores Extremos y LogPearson III locales, se ve igualada por la Pearson tipo III local e incluso disminuida por la función Gumbel.

Tal y como se desprende de los resultados analizados para cada estación de la región Norte, se puede decir que la diferencia entre el valor de un determinado cuantil estimado regionalmente y su estimación local surge no sólo por el empleo de funciones y parametrizaciones distintas, sino también por las variaciones entre ambas metodologías. Particularizar en cada sitio los valores de una curva regional adimensional construida mediante todas las observaciones de la región – caso del análisis regional - implica que cada estación va a estar condicionada por el resto. Sin embargo, con el análisis local son única y exclusivamente las propias observaciones las que determinan el resultado de la estimación. Así se explica que en ocasiones un cuantil estimado regionalmente sea mayor que el estimado a nivel local, y otras no.

Por otro lado, parece claro que en la región Norte el análisis regional presenta una incertidumbre menor que el local. Salvo para los casos de Velate, Zubiri y, en menor medida, Iraizoz, los límites de error al 90% constituyen un intervalo de confianza más reducido en el método regional. Se da la circunstancia de que son precisamente esas tres estaciones las que arrojan los mayores valores de discordancia con el resto de la región (ver tabla 3.7), con lo que su menor incertidumbre a nivel local queda explicada por el hecho de que, posiblemente, la curva regional adimensional – que aglutina las características medias de la región – no aproxima del todo bien a sus observaciones (más discordantes que el resto).

3.2.2.- REGIÓN SUR

ILUNDÁIN

Tabla 3.47. Cuantiles de precipitación en 24 h. de Ilundáin por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	48.22	43.12	52.93	47.52	42.71	51.60
5	60.98	54.64	67.94	59.98	53.41	67.36
10	68.62	60.97	78.59	68.23	59.82	78.34
20	75.45	66.15	89.42	76.14	65.63	88.81
50	83.72	72.47	101.74	86.38	73.06	102.21
100	89.58	76.01	115.20	94.05	78.54	112.16
200	95.19	79.56	126.69	101.69	83.94	122.04
500	102.32	84.25	138.76	102.32	89.01	132.43
1000	107.53	90.37	141.49	107.53	92.65	140.11

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	48.17	43.35	53.03	47.28	43.01	51.50
5	60.88	54.78	67.77	60.07	53.23	67.54
10	68.63	60.93	77.67	68.72	59.72	81.30
20	75.60	65.95	87.88	77.16	65.89	99.36
50	84.01	71.23	104.27	88.40	74.06	129.78
100	89.89	75.00	117.15	97.09	79.24	164.07
200	95.40	78.17	129.62	106.04	85.30	213.46
500	102.18	81.46	145.96	118.35	-	-
1000	106.96	83.74	159.09	128.07	-	-

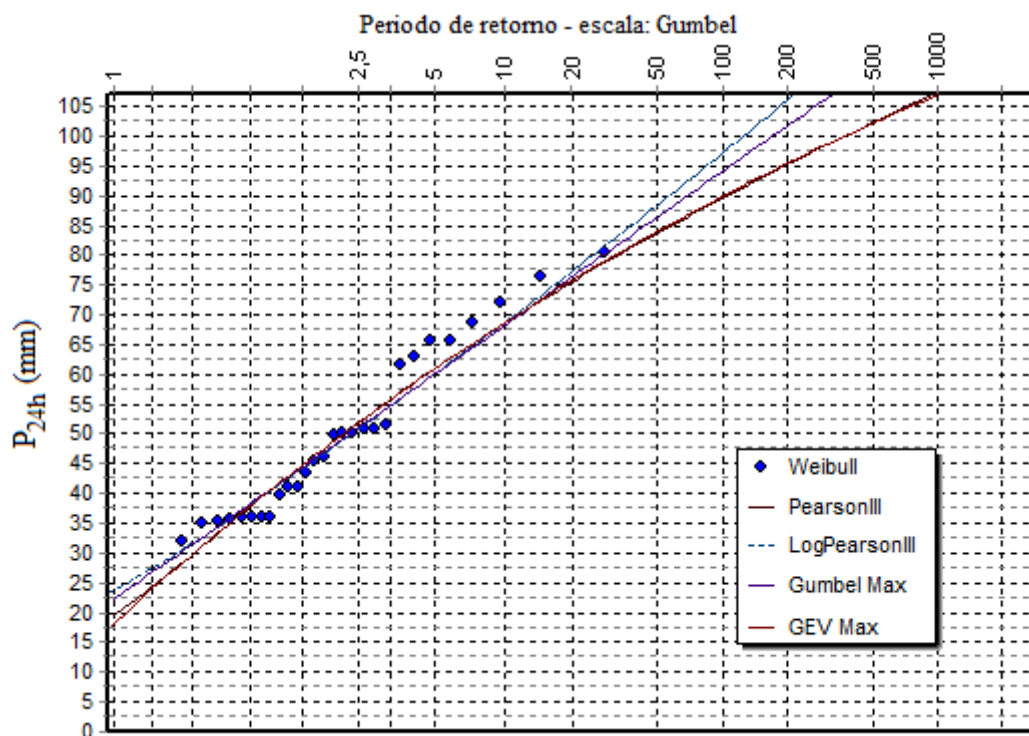


Fig. 3.33. Representación de los cuantiles de precipitación en 24 h. de Ilundáin por el método local.

MONREAL

Tabla 3.48. Cuantiles de precipitación en 24 h. de Monreal por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	45.79	39.39	52.27	46.25	39.77	51.73
5	62.80	53.73	73.11	62.47	53.55	72.53
10	73.87	61.80	88.84	73.22	61.89	87.02
20	84.21	68.64	104.90	83.52	69.54	100.84
50	97.19	76.30	125.89	96.86	79.09	118.63
100	106.65	81.19	140.88	106.85	86.09	131.92
200	115.88	85.54	157.72	116.81	92.85	145.24
500	127.82	91.10	180.19	129.95	100.23	169.11
1000	136.69	-	-	139.87	-	-

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	46.20	40.19	52.47	45.75	40.09	51.26
5	62.38	54.14	72.77	62.05	53.64	71.67
10	73.15	61.52	87.53	73.31	62.17	89.15
20	83.53	67.43	103.17	84.47	70.35	109.93
50	97.02	73.99	126.15	99.49	81.31	143.83
100	107.18	78.46	145.95	111.25	89.33	166.99
200	117.34	82.41	168.65	123.45	95.91	195.11
500	130.80	-	-	138.15	-	-
1000	141.03	-	-	153.51	-	-

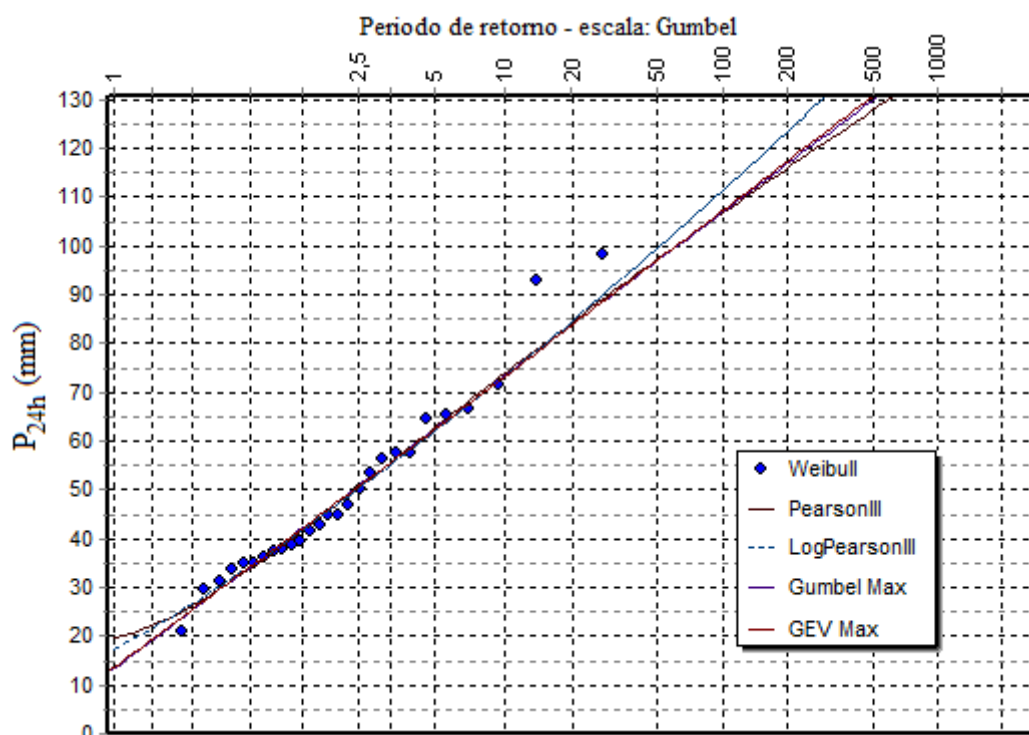


Fig. 3.34. Representación de los cuantiles de precipitación en 24 h. de Monreal por el método local.

NOÁIN

Tabla 3.49. Cuantiles de precipitación en 24 h. de Noáin por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	41.16	35.39	47.68	42.99	37.01	48.20
5	59.30	50.02	70.75	60.24	51.75	69.59
10	72.11	59.13	88.43	71.66	60.81	84.52
20	84.54	67.08	106.96	82.62	69.08	98.69
50	100.62	76.01	132.39	96.81	79.75	116.99
100	112.60	82.69	150.63	107.44	87.60	130.59
200	124.46	88.60	171.00	118.03	95.26	144.12
500	140.01	95.74	196.71	132.00	105.24	162.05
1000	151.69	-	-	142.56	112.85	175.74

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	42.35	37.12	48.25	41.48	37.00	46.75
5	59.01	51.34	69.66	57.94	50.20	67.25
10	70.78	59.02	85.38	70.49	59.39	85.81
20	82.64	65.56	102.13	83.82	68.29	112.02
50	98.91	73.01	126.69	103.16	80.46	158.60
100	111.79	78.32	146.92	119.37	90.11	204.53
200	125.27	83.88	168.23	137.16	100.18	258.95
500	144.05	-	-	163.43	-	-
1000	159.06	-	-	185.66	-	-

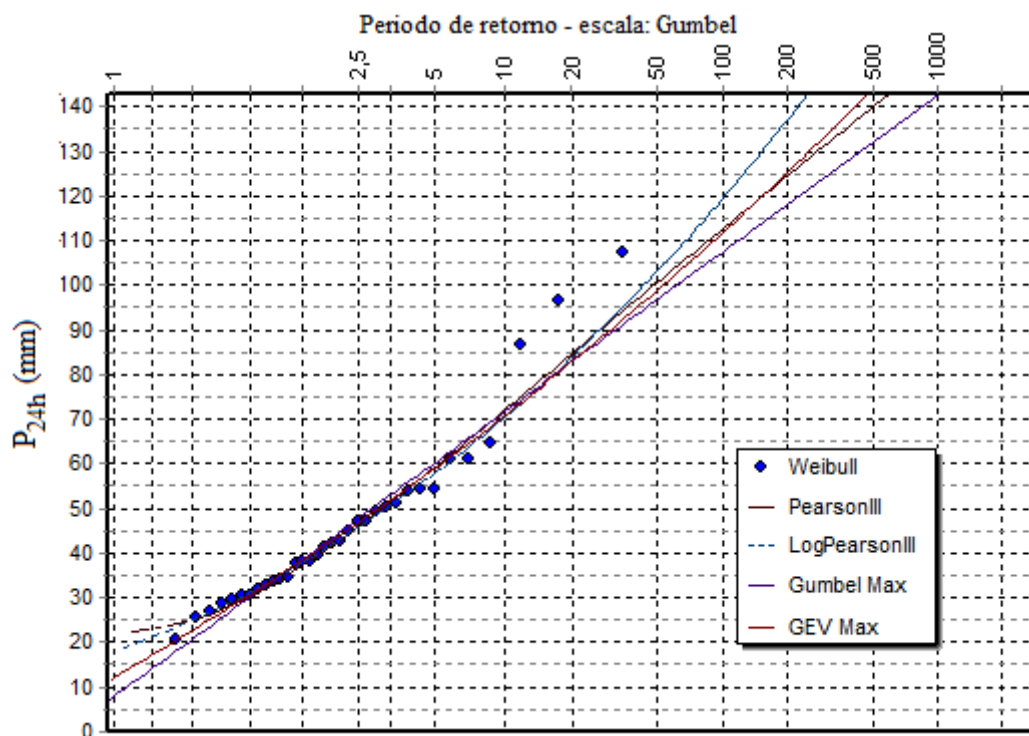


Fig. 3.35. Representación de los cuantiles de precipitación en 24 h. de Noáin por el método local.

OTAZU

Tabla 3.50. Cuantiles de precipitación en 24 h. de Otazu por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	45.35	40.97	49.59	44.73	40.53	48.31
5	57.42	51.86	63.42	56.50	50.62	62.93
10	64.66	57.95	73.13	64.28	56.77	73.04
20	71.15	62.75	84.36	71.75	62.43	82.65
50	79.01	68.60	97.89	81.42	69.56	94.99
100	84.58	72.52	110.21	88.67	74.87	104.19
200	89.92	76.04	125.34	95.89	80.17	113.38
500	96.71	80.74	136.58	105.42	90.11	120.54
1000	101.67	83.97	149.38	112.62	96.87	130.11

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	45.32	41.13	49.50	44.72	40.61	48.46
5	57.31	51.84	63.24	56.78	51.19	62.91
10	64.65	57.84	72.31	64.72	57.56	74.81
20	71.27	62.53	82.37	72.35	63.23	86.73
50	79.28	68.10	96.23	82.31	69.63	110.04
100	84.88	71.87	106.91	89.90	75.45	128.33
200	90.15	75.27	117.50	97.59	81.75	149.69
500	96.65	79.15	132.29	106.22	88.76	163.70
1000	101.24	81.35	144.09	115.63	-	-

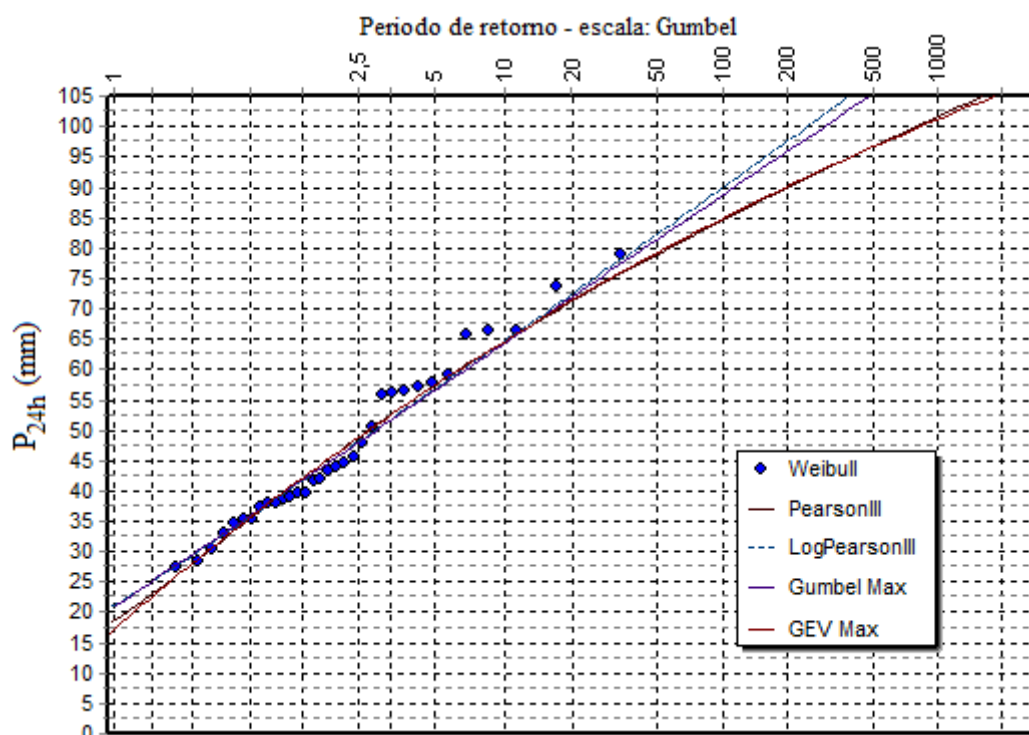


Fig. 3.36. Representación de los cuantiles de precipitación en 24 h. de Otazu por el método local.

PAMPLONA

Tabla 3.51. Cuantiles de precipitación en 24 h. de Pamplona por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	48.64	45.06	52.63	49.34	45.96	52.41
5	69.04	63.81	74.35	68.79	63.72	74.26
10	82.43	75.20	90.23	81.66	75.05	89.13
20	94.98	85.29	105.91	94.00	85.77	103.36
50	110.78	97.08	126.72	109.98	99.51	121.78
100	122.33	105.05	142.48	121.96	109.72	135.61
200	133.61	112.94	158.52	133.89	119.80	149.37
500	148.22	122.91	179.93	149.63	133.13	167.52
1000	159.09	130.05	196.12	161.53	143.23	181.23

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	49.22	45.99	52.71	48.07	45.23	51.46
5	68.57	63.78	74.03	67.37	62.54	72.97
10	81.51	74.79	88.83	81.35	74.28	89.92
20	94.04	84.72	104.81	95.65	85.60	108.69
50	110.40	97.03	127.86	115.59	100.12	138.91
100	122.77	105.89	147.00	131.68	110.35	166.64
200	135.20	114.36	168.70	148.79	121.44	200.52
500	151.75	124.73	199.00	173.20	134.15	272.51
1000	164.37	131.84	223.34	193.15	-	-

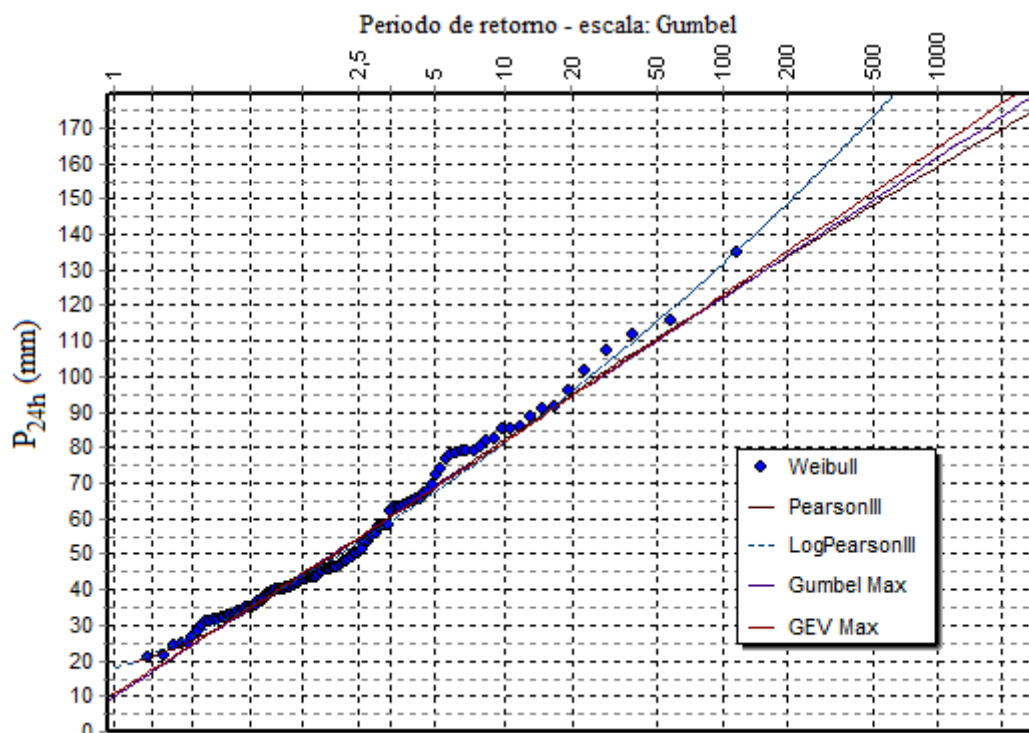


Fig. 3.37. Representación de los cuantiles de precipitación en 24 h. de Pamplona por el método local.

ZUAZU

Tabla 3.52. Cuantiles de precipitación en 24 h. de Zuazu por el método local.

T (años)	Pearson tipo III			Gumbel		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	45.62	37.90	53.42	47.25	39.39	53.81
5	63.44	52.37	78.38	64.19	53.35	76.13
10	75.89	60.87	97.74	75.40	61.85	91.88
20	87.92	67.87	117.88	86.16	69.52	106.91
50	103.42	75.94	143.66	100.09	79.25	125.96
100	114.94	81.63	162.36	110.52	86.46	140.01
200	126.33	87.43	180.18	120.92	93.60	154.06
500	141.23	93.89	207.98	134.63	102.83	172.66
1000	152.41	-	-	145.00	-	-

T (años)	General Valores Extremos			LogPearson III		
	P _{24h} (mm)	Lím 0.05	Lím 0.95	P _{24h} (mm)	Lím 0.05	Lím 0.95
2	46.69	39.71	54.11	45.87	40.17	52.53
5	63.11	53.38	77.06	62.27	52.29	74.72
10	74.64	60.28	93.49	74.65	60.14	95.61
20	86.21	66.11	110.25	87.71	67.49	123.90
50	101.97	72.29	135.50	106.52	78.21	170.67
100	114.38	76.10	157.75	122.18	86.75	216.50
200	127.30	79.69	182.60	139.27	96.21	292.72
500	145.22	-	-	164.38	-	-
1000	159.45	-	-	185.50	-	-

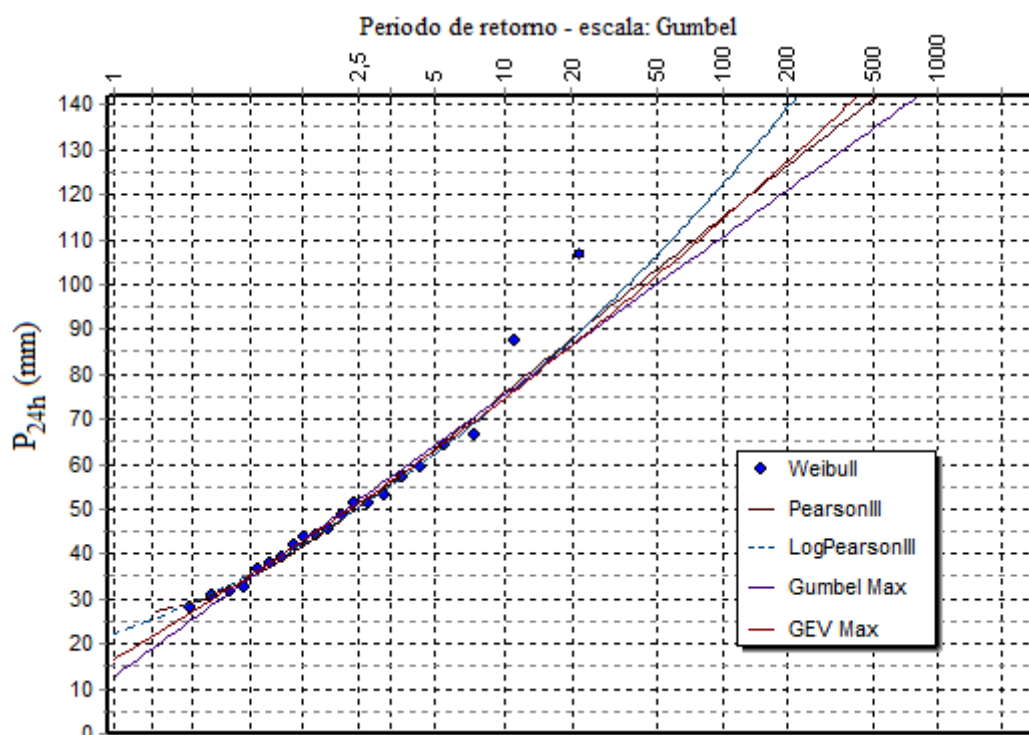


Fig. 3.38. Representación de los cuantiles de precipitación en 24 h. de Zuazu por el método local.

Se analizan los resultados obtenidos para cada estación de la región Sur:

- Ilundáin. Se aprecia cómo los cuantiles obtenidos por análisis regional resultan superiores a los estimados localmente. Por otro lado, la incertidumbre de la estimación regional de la distribución Pearson tipo III es menor que la local de LogPearson III, pero mayor que la de Pearson III, Gumbel y General de Valores Extremos empleadas localmente.
- Monreal. Para los periodos de retorno de 2, 5 y 10 años las estimaciones regionales y locales resultan parecidas. A partir del periodo de retorno de 20 años, los cuantiles estimados regionalmente superan a los obtenidos por el método local. En este caso, la incertidumbre del análisis regional resulta menor que la de las estimaciones locales, con la salvedad de los periodos de retorno de 200 y 500 años de la distribución Gumbel.
- Noáin. Pese a que para los periodos de retorno de 2, 5 y 10 años las estimaciones son parecidas, a partir del periodo de retorno de 20 años los cuantiles estimados por el método regional son inferiores a los obtenidos localmente. Al igual que en Monreal, la incertidumbre del análisis regional resulta menor que la de las estimaciones locales salvo para los cuantiles más extremos (periodos de retorno de 500 y 1000 años) de la distribución Gumbel.
- Otazu. Se observa cómo, salvo para el periodo de retorno de 2 años, los cuantiles obtenidos por análisis regional resultan notablemente superiores a los estimados localmente. En esta ocasión la incertidumbre de la estimación regional de la distribución Pearson tipo III tan sólo es inferior a la LogPearson III local, siendo superior a la obtenida mediante las distribuciones Pearson III, Gumbel y General de Valores Extremos locales.
- Pamplona. Los cuantiles estimados tanto por el método regional como por el local resultan prácticamente similares, si bien los obtenidos mediante la distribución LogPearson III local se disparan en los periodos de retorno más extremos (500 y 1000 años). En lo que respecta a la incertidumbre, ésta resulta algo menor en el análisis regional mediante la distribución Pearson III, aunque la aplicación local de Gumbel presenta la mayor fiabilidad, sobretudo en los cuantiles más extremos.
- Zuazu. Se aprecia cómo los cuantiles del análisis regional son muy parecidos a los obtenidos localmente, salvo para el caso de la distribución LogPearson III, donde se disparan. Por otro lado, la estimación regional conlleva una incertidumbre claramente inferior a la local.

Nuevamente quedan de manifiesto las diferencias entre el valor de un determinado cuantil estimado regionalmente y su estimación local como consecuencia tanto del empleo de funciones y parametrizaciones distintas como por las variaciones entre ambas metodologías.

Por otro lado, tal y como ocurría en la región Norte, en la región Sur el análisis regional presenta una incertidumbre inferior a la del local, aunque con diferencias menos marcadas probablemente como consecuencia del menor número de estaciones. Salvo para los casos de Ilundáin y Otazu, los límites de error al 90% constituyen un intervalo de confianza más reducido en el método regional. Al igual que antes, son precisamente esas dos estaciones las que arrojan los mayores valores de discordancia con el resto de la región Sur (ver tabla 3.7), con lo que su menor incertidumbre a nivel local viene a confirmar el hecho de que, posiblemente, la curva regional adimensional – que aglutina las características medias de la región – no aproxima del todo bien a sus observaciones (más discordantes que el resto).

La comparativa entre la estimación regional y local en la cuenca del río Arga - atendiendo tanto a los resultados de la región Norte como de la región Sur - da una idea sobre la conveniencia de optar por uno u otro método.

Si bien el análisis regional ha conseguido alcanzar resultados de menor incertidumbre, no hay que olvidar que, aun habiéndose implementado en los cálculos, la existencia de dependencia entre las estaciones puede presentarse como un problema en muchas situaciones, pues el método la asume como inexistente. En este sentido, parece que la metodología del análisis regional requeriría de un ajuste y redefinición de la medida de heterogeneidad H que tuviera en cuenta la dependencia interestacional. Pese a ello, el caso concreto de la cuenca del Arga no tiene por qué verse afectado salvo por un ligero aumento en el error cuadrático medio de las estimaciones, dado que tanto la región Norte como la Sur son moderadamente heterogéneas. Abarcar una zona de estudio más amplia cuyas estaciones estuvieran más distanciadas garantizaría la independencia, pero al mismo tiempo dificultaría la regionalización y las estimaciones de cada estación podrían verse condicionadas por las observaciones de otras muy diferentes a las suyas propias.

Por otro lado, parece claro que, a excepción de Pamplona, las escasas longitudes de los registros de las estaciones hacen menos recomendable el empleo del método local. Se ha comprobado cómo en algunas ocasiones la incertidumbre de éste resulta inferior a la del método regional, sobretudo para casos concretos de la distribución Gumbel, pero ello es debido únicamente a su adecuado ajuste a los valores observados; de hecho, en un registro reducido de años de cierta estación han podido no quedar reflejados muchos eventos extremos ocurridos y registrados en las estaciones cercanas.

Por todo ello, y sin olvidar la ventaja que supone caracterizar toda una región y no sólo una estación puntual, el análisis regional parece la opción más aconsejable para la estimación de las frecuencias de las precipitaciones en 24 h. en la cuenca del río Arga.

4.- CONCLUSIONES

La aplicación del análisis regional de frecuencia de las precipitaciones máximas en 24 h. en la cuenca del río Arga, con el empleo del método del índice de avenida basado en los *L*-momentos, permite extraer las siguientes conclusiones:

- La transición climática, biogeográfica y física que se da a lo largo de la cuenca, junto con los registros de las 15 estaciones meteorológicas manuales analizadas, reflejan que toda ella no puede ser considerada como una región homogénea. El estudio en su regionalización resulta en dos zonas diferenciadas: la Norte, de 9 estaciones (Belzunce, Erro, Espinal, Eugui, Iraizoz, Iroz, Olagüe, Velate y Zubiri), correspondiente a la cabecera de la cuenca, y la Sur, con 6 estaciones (Ilundáin, Monreal, Noáin, Otazu, Pamplona y Zuazu), correspondiente a la comarca de Pamplona.
- El estudio de heterogeneidad de las regiones Norte y Sur revela que ambas son homogéneas si se consideran cumplidas las suposiciones que asume el método del índice de avenida. No obstante, la evidente existencia de dependencia entre sus estaciones conduce a calificarlas como moderadamente heterogéneas.
- Son tres las distribuciones triparamétricas que aportan un ajuste aceptable a la región Norte: General de Valores Extremos (GEV), LogNormal 3 (LN3), y Pearson tipo III (PE3). Entre ellas apenas existen diferencias. No obstante, la distribución Pearson tipo III (PE3), además de conseguir el mejor ajuste a las observaciones, arroja el menor Error Cuadrático Medio relativo para las estimaciones. Por ello se opta por su elección para la zona Norte de cara a conseguir el modelo más robusto.
- En la región Sur, son tres las distribuciones triparamétricas que aportan un ajuste adecuado: LogNormal 3 (LN3), Pearson tipo III (PE3) y General de Pareto (GPA). En este caso, sus notables diferencias a partir del periodo de retorno de 50 años llevan a considerar también la función tetraparamétrica Kappa (KAP). No obstante, nuevamente la distribución Pearson tipo III (PE3) arroja el menor Error Cuadrático Medio relativo para las estimaciones, además del mejor ajuste a las observaciones, por lo que se opta por su elección para la región Sur de cara a un modelo robusto.
- La particularización de la distribución Pearson tipo III (PE3) de cada región en cada una de las estaciones que la integran confirma que los cuantiles de precipitación en 24 h. son mayores en aquellas estaciones que presentan mayor factor de escala (índice de avenida): Eugui y Velate en la región Norte, y Pamplona y Zuazu en la región Sur.

Por otro lado, la comparación del análisis regional de frecuencia con los métodos clásicos de análisis local aplicados en las precipitaciones máximas en 24 h. en la cuenca del río Arga, permite extraer las siguientes conclusiones:

- La diferencia entre el valor de un determinado cuantil estimado regionalmente y su estimación local surge por el empleo de funciones y parametrizaciones distintas, pero también por las variaciones entre ambas metodologías: emplear todas las observaciones de la región para estimar los cuantiles de cada estación (análisis regional), o utilizar únicamente las propias observaciones de la estación (análisis local).
- En la región Norte, el análisis regional presenta una incertidumbre menor que el local, salvo para las estaciones de Iraizoz, Velate y Zubiri.
- Igualmente, en la región Sur, salvo para las estaciones de Ilundáin y Otazu, el análisis regional presenta una incertidumbre menor que el local, aunque con diferencias menos acusadas como consecuencia del menor número de estaciones en la región.
- La moderada heterogeneidad de las regiones Norte y Sur definidas hace aconsejable la estimación de cuantiles mediante el análisis regional de frecuencias.
- Las escasas longitudes de los registros de observaciones de las estaciones hacen menos recomendable el empleo del método clásico local.

5.- REFERENCIAS

- Acreman, M.C., Sinclair, C.D. (1986). Classification of drainage basins according to their physical characteristics: An application for flood frequency análisis in Scotland. *Journal of Hydrology* 84, 365-380.
- Ahmad, M.I., Sinclair, C.D., Werritty, A. (1988). Log-logistic flood frequency analysis. *Journal of Hydrology* 98, 215-224.
- Álvarez, M., Puertas, J., Soto, B., Díaz-Fierros, F. (1999). Análisis regional de las precipitaciones máximas en Galicia mediante el método del índice de avenida. *Ingeniería del Agua* 6 (4), 283-290.
- Barnett, V., Lewis, T. (1994). *Outliers in statistical data*, 3rd ed. Wiley, Chichester, UK.
- Beable, M.E., McKerchar, A.I. (1982). *Regional flood estimation in New Zealand*. Water and Soil Technical Publication 20, Ministry of Works and Development, Wellington, N.Z.
- Boes, D.C., Heo, J.-H., Salas, J.D. (1989). Regional flood quantile estimation for a Weibull model. *Water Resources Research* 25, 979-990.
- Boughton, W.C. (1980). A frequency distribution for annual floods. *Water Resources Research* 16, 347-354.
- Buishand, T.A. (1989). Statistics of extremes in climatology. *Statistica Neerlandica* 43, 1-30.
- Chowdhury, J.U., Stedinger, J.R., Lu, L.-H. (1991). Goodness-of-fit tests for regional generalized extreme value flood distributions. *Water Resources Research* 27, 1765-1776.
- Dalrymple, T. (1960). *Flood frequency analysis*. Water Supply Paper 1543-A, USGS, Reston, VA, USA.
- Damazio, J.M, Kelman, J. (1986). Use of historical data in flood-frequency analysis. *International Symposium on Flood Frequency and Risk Analysis*, Louisiana State University, Baton Rouge, La., May 1986.
- Documento Técnico para la Participación Pública en la Cuenca del Arga. Foro del Agua de Navarra (2007). Departamento de Medio Ambiente, Ordenación del Territorio y Vivienda. Gobierno de Navarra.
- Ferrer, J., Ardiles, L., (1994). Análisis estadístico de las series anuales de máximas lluvias diarias en España. *Ingeniería Civil* 95, 87-100.
- Fill, H.D., Stedinger, J.R., (1995). Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple's test. *Journal of Hydrology* 166, 81-105.

- García, A.P. (2007). Análisis multifractal de series de datos pluviométricos en Andalucía. Escuela Técnica Superior de Ingenieros Agrónomos y de Montes. Universidad de Córdoba.
- Gómez, P., Sánchez, R., Sánchez, C. (1996). As precipitacións. As Augas de Galicia. Consello da Cultura Galega. Ponencia de Patrimonio Natural. Santiago de Compostela, 43-105.
- Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research* 15 (5), 1049–1054.
- Gumbel, E.J. (1958). *Statistics of extremes*. Columbia University Press, New York.
- Guttman, N.B. (1993). The use of L-moments in the determination of regional precipitation climates. *Journal of Climate* 6, 2309-2325.
- Guttman, N.B., Hosking, J.R.M., Wallis, J.R. (1993). Regional precipitation quantile values for the continental US computed from L-moments. *Journal of Climate* 6, 2326–2340.
- Hosking, J.R.M. (1990). L-moments: Analysis and estimación of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, 52, 105-124.
- Hosking, J.R.M. (1996). Fortran routines for use with the method of L-moments, Version 3. Research Report RC 20525, IBM Research Division, Yorktown, Heights, N.Y.
- Hosking, J.R.M. (2009). L-moments. R package, version 1.5. URL: <http://CRAN.R-project.org/package=lmom>.
- Hosking, J.R.M. (2010). Regional frequency analysis using L-moments. R package, version 2.3. URL: <http://CRAN.R-project.org/package=lmomRFA>.
- Hosking, J.R.M., Wallis, J.R. (1986). The value of historical data in flood frequency analysis. *Water Resources Research* 22, 1606-1612.
- Hosking, J.R.M., Wallis, J.R. (1987). Correlation and dependence between annual maximum flood series. Research Report RC12822, IBM Research Division, Yorktown Heights, N.Y.
- Hosking, J.R.M., Wallis, J.R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research* 29 (2), 271–281.
- Hosking, J.R.M., Wallis, J.R. (1997). *Regional Frequency Analysis: An Approach based on L-moments*. Cambridge University Press, Cambridge, UK.

- Infraestructura de Datos Espaciales de Navarra IDENA. Gobierno de Navarra. URL: <http://idena.navarra.es/>.
- ITIA Research Team (2010). Athens, Greece. URL: <http://hydrognomon.org/>.
- Jiménez, A., Mediero, L. (2009). Caracterización del comportamiento estadístico de los caudales máximos anuales y estacionales de los ríos de la España peninsular. Análisis y propuesta de procedimientos para su determinación. Centro de Estudios Hidrográficos del CEDEX. Jornadas sobre ingeniería del agua.
- Kendall, M.G. (1975). Rank Correlation Methods. Griffin, London, UK.
- Kotz, S., Johnson, N.L., Read, C.B. (1988). Encyclopedia of Statistical Sciences, vol. 9. Wiley, New York.
- Landwehr, J.M., Matalas, N.C., Wallis, J.R. (1979). Probability-weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resources Research 15 (1979), 983-4).
- Laursen, E.M. (1983). Comentario en “Paleohydrology of southwestern Texas” por R.C. Kochel, V.R. Baker y P.C. Patton. Water Resources Research 19, 1339.
- Lu, L.-H., Stedinger, J.R. (1992). Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. Journal of Hydrology 138, 223-245.
- Madsen, H., Pearson, C.P., Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 2. Regional modeling. Water Resources Research 33 (4), 759-769.
- McMahon, T.A., Srikanthan, R. (1982). Log Pearson type 3 distribution effect of dependence, distribution parameters and sample size on peak annual flood estimates. Journal of Hydrology 52, 149-159.
- Meteorología y Climatología de Navarra. Departamentos de Desarrollo Rural e Innovación. Gobierno de Navarra. URL: <http://meteo.navarra.es/>.
- MMA (1998). Las precipitaciones máximas en 24 horas y sus períodos de retorno en España. Un estudio por regiones. Ministerio de Medio Ambiente. INM.
- MOPTMA (1995). Valores normales estadísticos de estaciones principales (1961-1990). Ministerio de Obras Públicas Transportes y Medio Ambiente. INM, publicación K-68.
- Munro, A.H., Wixley, R.A.J. (1970). Estimators based on order statistics of small samples from a three-parameter lognormal distribution. Journal of the American Statistical Association 65, 212 – 225.

- Norbiato, D., Borga, M., Sangati, M., Zanon, F. (2007). Regional frequency analysis of extreme precipitation in the eastern Italian Alps and the August 29, 2003 flash flood. *Journal of Hydrology* 345, 149–166.
- Pearson, C.P., (1991). New Zealand regional flood frequency analysis using L-moments. *Journal of Hydrology (New Zealand)* 30, 53-64.
- Pearson, C.P., (1993). Application of L-moments to maximum river flows. *The New Zealand Statistician* 28, 2-10.
- Pearson, C.P., McKerchar, A.I., Woods, R.A. (1991). Regional flood frequency análisis of western Australian data using L-moments. *International Hydrology and Water Resources Symposium, Perth, Australia*, 631-632.
- Pilion, P.J., Admowski, K., Alila, Y. (1991). Regional analysis of annual maxima precipitation using L-moments. *Atmosphere Research* 27, 81–92.
- Rosbjerg, D., Madsen, H. (1995). Uncertainty measures of regional flood frequency estimators. *Journal of Hydrology* 167, 209-224.
- R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: <http://www.R-project.org>.
- Saenz de Ormijana, F., Hidalgo, F.J., Santa Pérez, A. (1991). Estimación de las precipitaciones máximas mediante el método regional del índice de avenida. *Revista de Obras Públicas*, Feb., 9-22.
- Schaefer, M.G. (1990). Regional analysis of precipitation annual maximum in Washington State. *Water Resources Research* 26, 119–131.
- Sistema de Información Territorial de Navarra SITNA. Gobierno de Navarra. URL: <http://sitna.navarra.es/>.
- Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E. (1992) Frequency analysis of extreme events. En *Handbook of Hydrology*, editado por D.R. Maidment, Capítulo 18. McGraw-Hill, New York.
- Vogel, R.M., Thomas, W.O., McMahon, T.A. (1993). Flood-flow frequency model selection in southwestern United States. *Journal of Water Resources Planning and Management* 119 (3), 353-366.
- Wallis, J.R. (1993). Regional frequency studies using L-moments. En *Concise Encyclopedia of Environmental Systems*, editada por P.C. Young, pp. 468-476. Pergamon, Oxford.
- Wiltshire, S.E. (1986a). Regional flood frequency analysis I: Homogeneity statistics. *Hydrological Sciences Journal* 31, 321-333.

- Wiltshire, S.E. (1986b). Regional flood frequency analysis II: Multivariate classification of drainage basins in Britain. *Hydrological Sciences Journal* 31, 335-346.
- Yang, T., Shao, Q., Hao, Z., Chen, X., Zhang, Z., Xu, Ch., Sun, L. (2010). Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin, China. *Journal of Hydrology* 380, 386–405.

ANEXO I: DISTRIBUCIONES Y SUS L-MOMENTOS

I.1.- DISTRIBUCIÓN GENERAL DE PARETO

I.1.1.- DEFINICIÓN

Parámetros (3): ξ (localización), α (escala), k (forma).

Rango de x : $\xi \leq x \leq \xi + \alpha/k$ para $k > 0$; $\xi \leq x < \infty$ para $k \leq 0$.

$$f(x) = \alpha^{-1} e^{-(1-k)y}, \quad y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases} \quad (\text{I.1})$$

$$F(x) = 1 - e^{-y} \quad (\text{I.2})$$

$$x(F) = \begin{cases} \xi + \alpha\{1 - (1-F)^k\}/k, & k \neq 0 \\ \xi - \alpha \log(1-F), & k = 0 \end{cases} \quad (\text{I.3})$$

Casos especiales: $k = 0$ es la distribución exponencial; $k = 1$ es la distribución uniforme en el intervalo $\xi \leq x \leq \xi + \alpha$.

I.1.2.- L-MOMENTOS

Los L-momentos se definen para $k > -1$.

$$\lambda_1 = \xi + \alpha/(1+k) \quad (\text{I.4})$$

$$\lambda_2 = \alpha/\{(1+k)(2+k)\} \quad (\text{I.5})$$

$$\tau_3 = (1-k)/(3+k) \quad (\text{I.6})$$

$$\tau_4 = (1-k)(2-k)/\{(3+k)(4+k)\} \quad (\text{I.7})$$

La relación entre τ_3 y τ_4 viene dada por

$$\tau_4 = \frac{\tau_3(1+5\tau_3)}{5+\tau_3} \quad (\text{I.8})$$

I.1.3.- PARÁMETROS

Si ξ es conocido, los dos parámetros α y k vienen dados por

$$k = (\lambda_1 - \xi)/\lambda_2 - 2, \quad \alpha = (1+k)(\lambda_1 - \xi) \quad (\text{I.9})$$

Si ξ es desconocido, los tres parámetros vienen dados por

$$k = (1-3\tau_3)/(1+\tau_3), \quad \alpha = (1+k)(2+k)\lambda_2, \quad \xi = \lambda_1 - (2+k)\lambda_2 \quad (\text{I.10})$$

I.2.- DISTRIBUCIÓN GENERAL DE VALORES EXTREMOS

I.2.1.- DEFINICIÓN

Parámetros (3): ξ (localización), α (escala), k (forma).

Rango de x : $-\infty < x \leq \xi + \alpha/k$ para $k > 0$; $-\infty < x < \infty$ para $k = 0$; $\xi + \alpha/k \leq x < \infty$ para $k < 0$.

$$f(x) = \alpha^{-1} e^{-(1-k)y}, \quad y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases} \quad (\text{I.11})$$

$$F(x) = e^{-e^{-y}} \quad (\text{I.12})$$

$$x(F) = \begin{cases} \xi + \alpha\{1 - (-\log F)^k\}/k, & k \neq 0 \\ \xi - \alpha \log(-\log F), & k = 0 \end{cases} \quad (\text{I.13})$$

Casos especiales: $k = 0$ es la distribución Gumbel; $k = 1$ es una distribución exponencial revertida; esto es, $1 - F(-x)$ es la función de distribución acumulada de una distribución exponencial.

Las distribuciones de Valores Extremos se dividen en tres tipos, caracterizados por las siguientes funciones de distribución acumulada:

$$\text{Tipo I: } F(x) = \exp(e^{-x}), \quad -\infty < x < \infty \quad (\text{I.14})$$

$$\text{Tipo II: } F(x) = \exp(-x^{-\delta}), \quad 0 \leq x < \infty \quad (\text{I.15})$$

$$\text{Tipo III: } F(x) = \exp(-|x|^\delta), \quad -\infty < x \leq 0 \quad (\text{I.16})$$

La distribución General de Valores Extremos recoge todos estos tipos, I, II y III, que corresponden a $k = 0$, $k < 0$, y $k > 0$, respectivamente. La distribución Weibull, definida por

$$F(x) = 1 - \exp[-\{(x - \zeta)/\beta\}^\delta], \quad \zeta \leq x < \infty \quad (\text{I.17})$$

es una distribución General de Valores Extremos revertida, con parámetros

$$k = 1/\delta, \quad \alpha = \beta/\delta, \quad \xi = \zeta - \beta \quad (\text{I.18})$$

I.2.2.- L-MOMENTOS

Los L-momentos están definidos para $k > -1$.

$$\lambda_1 = \xi + \alpha\{1 - \Gamma(1+k)\}/k \quad (\text{I.19})$$

$$\lambda_2 = \alpha(1 - 2^{-k})\Gamma(1+k)/k \quad (\text{I.20})$$

$$\tau_3 = 2(1 - 3^{-k})/(1 - 2^{-k}) - 3 \quad (\text{I.21})$$

$$\tau_4 = \{5(1 - 4^{-k}) - 10(1 - 3^{-k}) + 6(1 - 2^{-k})\}/(1 - 2^{-k}) \quad (\text{I.22})$$

Aquí $\Gamma(\cdot)$ denota la función Gamma

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (I.23)$$

I.2.3.- PARÁMETROS

El valor de k se estima resolviendo la expresión (I.21) para k . No es posible obtener una solución explícita, pero la aproximación siguiente tiene una precisión mayor que 9×10^{-4} para $-0.5 \leq \tau_3 \leq 0.5$:

$$k \approx 7.8590c + 2.9554c^2, \quad c = \frac{2}{3 + \tau_3} - \frac{\log 2}{\log 3} \quad (I.24)$$

Los otros parámetros vienen dados por

$$\alpha = \frac{\lambda_2 k}{(1 - 2^{-k})\Gamma(1+k)}, \quad \xi = \lambda_1 - \alpha \{1 - \Gamma(1+k)\} / k. \quad (I.25)$$

I.3.- DISTRIBUCIÓN LOGÍSTICA GENERALIZADA

I.3.1.- DEFINICIÓN

Parámetros (3): ξ (localización), α (escala), k (forma).

Rango de x : $-\infty < x \leq \xi + \alpha/k$ si $k > 0$; $-\infty < x < \infty$ si $k = 0$; $\xi + \alpha/k \leq x < \infty$ si $k < 0$.

$$f(x) = \frac{\alpha^{-1} e^{-(1-k)y}}{(1 + e^{-y})^2}, \quad y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases} \quad (I.26)$$

$$F(x) = 1/(1 + e^{-y}) \quad (I.27)$$

$$x(F) = \begin{cases} \xi + \alpha[1 - \{(1 - F)/F\}^k] / k, & k \neq 0 \\ \xi - \alpha \log\{(1 - F)/F\}, & k = 0 \end{cases} \quad (I.28)$$

Casos especiales: $k = 0$ es la distribución Logística.

Esta generalización de la distribución Logística difiere de otras definidas en la literatura. Se trata de una versión reparametrizada de la distribución LogLogística de Ahmad et al. (1988). Se ha adoptado este nombre para reflejar la similitud de la distribución con la General de Pareto y con la General de Valores Extremos.

I.3.2.- L-MOMENTOS

Los L-momentos vienen definidos para $-1 < k < 1$.

$$\lambda_1 = \xi + \alpha(1/k - \pi / \sin k\pi) \quad (I.29)$$

$$\lambda_2 = \alpha k \pi / \sin k\pi \quad (I.30)$$

$$\tau_3 = -k \quad (\text{I.31})$$

$$\tau_4 = (1 + 5k^2)/6 \quad (\text{I.32})$$

I.3.3.- PARÁMETROS

$$k = -\tau_3, \quad \alpha = \frac{\lambda_2 \sin k\pi}{k\pi}, \quad \xi = \lambda_1 - \alpha \left(\frac{1}{k} - \frac{\pi}{\sin k\pi} \right) \quad (\text{I.33})$$

I.4.- DISTRIBUCIÓN LOGNORMAL

I.4.1.- DEFINICIÓN

Parámetros (3): ξ (localización), α (escala), k (forma).

Rango de x : $-\infty < x \leq \xi + \alpha/k$ para $k > 0$; $-\infty < x < \infty$ para $k = 0$; $\xi + \alpha/k \leq x < \infty$ para $k < 0$.

$$f(x) = \frac{e^{ky - y^2/2}}{\alpha\sqrt{2\pi}}, \quad y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases} \quad (\text{I.34})$$

$$F(x) = \Phi(y) \quad (\text{I.35})$$

$x(F)$ no tiene forma analítica explícita

Aquí Φ es la función de distribución acumulada de una distribución Normal estándar.

Caso especial: $k = 0$ es la distribución Normal con parámetros ξ y α .

La distribución LogNormal se define habitualmente por

$$F(x) = \Phi[\{\log(x - \zeta) - \mu\}/\sigma], \quad \zeta \leq x < \infty \quad (\text{I.36})$$

La reparametrización de la distribución LogNormal en términos de ξ , α y k es una pequeña modificación de la parametrización de Munro y Wixley (1970). Posee varias ventajas respecto a la parametrización habitual con μ , σ y ζ :

- dentro de una única distribución incluye la distribución LogNormal con asimetría positiva y límite inferior ($k < 0$), y la distribución LogNormal con asimetría negativa y límite superior ($k > 0$);
- incluye la distribución Normal como un caso especial ($k = 0$) en lugar de como un límite inalcanzable.
- muestra la similitud entre la estructura de la distribución LogNormal, la General de Pareto y la General de Valores Extremos;
- sus parámetros son más significativos y más estables a la hora de estimar sus valores que los de la parametrización estándar, particularmente cuando la asimetría es próxima a cero.

En esta parametrización, la distribución LogNormal es la de una variable aleatoria X relacionada con otra variable aleatoria Z de distribución Normal estándar de media 0 varianza 1, mediante

$$X = \begin{cases} \xi + \alpha(1 - e^{-kZ})/k, & k \neq 0 \\ \xi + \alpha Z, & k = 0 \end{cases} \quad (\text{I.37})$$

La parametrización estándar (I.36) puede ser obtenida a partir de la presentada aquí mediante

$$k = -\sigma, \quad \alpha = \sigma e^{\mu}, \quad \xi = \varsigma + e^{\mu} \quad (\text{I.38})$$

y se podrían obtener expresiones para ella a partir de las mostradas a continuación.

I.4.2.- L-MOMENTOS

Los L-momentos se definen para cualquier valor de k .

$$\lambda_1 = \xi + \alpha(1 - e^{k^2/2})/k \quad (\text{I.39})$$

$$\lambda_2 = \frac{\alpha}{k} e^{k^2/2} \{1 - 2\Phi(-k/\sqrt{2})\} \quad (\text{I.40})$$

No hay expresiones simples para los ratios de L -momentos. Son función únicamente de k , y pueden ser computados por integración numérica, como en Hosking (1996). Alternativamente, se pueden usar aproximaciones basadas en funciones racionales. Las siguientes aproximaciones para τ_3 y τ_4 tienen mayor precisión que 2×10^{-7} y 5×10^{-7} , respectivamente, para $|k| \leq 4$, lo que corresponde a $|\tau_3| \leq 0.99$ y $\tau_4 \leq 0.98$:

$$\tau_3 \approx -k \frac{A_0 + A_1 k^2 + A_2 k^4 + A_3 k^6}{1 + B_1 k^2 + B_2 k^4 + B_3 k^6} \quad (\text{I.41})$$

$$\tau_4 \approx \tau_4^0 + k^2 \frac{C_0 + C_1 k^2 + C_2 k^4 + C_3 k^6}{1 + D_1 k^2 + D_2 k^4 + D_3 k^6} \quad (\text{I.42})$$

Los coeficientes usados en las aproximaciones vienen dados en la tabla I.1.

I.4.3.- PARÁMETROS

El parámetro de forma, k , es función solamente de τ_3 . No hay solución explícita posible, pero la siguiente aproximación tiene una precisión relativa mayor que 2.5×10^{-6} para $|\tau_3| \leq 0.94$, lo que corresponde a $|k| \leq 3$:

$$k \approx -\tau_3 \frac{E_0 + E_1 \tau_3^2 + E_2 \tau_3^4 + E_3 \tau_3^6}{1 + F_1 \tau_3^2 + F_2 \tau_3^4 + F_3 \tau_3^6} \quad (\text{I.43})$$

Los coeficientes utilizados en la aproximación son los mostrados en la tabla I.1. El resto de parámetros vienen pues dados por

$$\alpha = \frac{\lambda_2 k e^{-k^2/2}}{1 - 2\Phi(-k/\sqrt{2})}, \quad \xi = \lambda_1 - \frac{\alpha}{k} (1 - e^{k^2/2}) \quad (\text{I.44})$$

Tabla I.1. Coeficientes de las aproximaciones (I.41) hasta (I.43).

$\tau_4^0 = 1.2260172 \times 10^{-1}$		
$A_0 = 4.8860251 \times 10^{-1}$	$C_0 = 1.8756590 \times 10^{-1}$	$E_0 = 2.0466534$
$A_1 = 4.4493076 \times 10^{-3}$	$C_1 = -2.5352147 \times 10^{-3}$	$E_1 = -3.6544371$
$A_2 = 8.8027039 \times 10^{-4}$	$C_2 = 2.6995102 \times 10^{-4}$	$E_2 = 1.8396733$
$A_3 = 1.1507084 \times 10^{-6}$	$C_3 = -1.8446680 \times 10^{-6}$	$E_3 = -0.20360244$
$B_1 = 6.4662924 \times 10^{-2}$	$D_1 = 8.2325617 \times 10^{-2}$	$F_1 = -2.0182173$
$B_2 = 3.3090406 \times 10^{-3}$	$D_2 = 4.2681448 \times 10^{-3}$	$F_2 = 1.2420401$
$B_3 = 7.4290680 \times 10^{-5}$	$D_3 = 1.1653690 \times 10^{-4}$	$F_3 = -0.21741801$

I.5.- DISTRIBUCIÓN DE PEARSON TIPO III

I.5.1.- DEFINICIÓN

Parámetros (3): μ (localización), σ (escala), γ (forma).

Si $\gamma \neq 0$, sea $\alpha = 4/\gamma^2$, $\beta = 1/2 \sigma|\gamma|$, y $\xi = \mu - 2\sigma/\gamma$. Si $\gamma > 0$, el rango de x es $\xi \leq x < \infty$, y

$$f(x) = \frac{(x - \xi)^{\alpha-1} e^{-(x-\xi)/\beta}}{\beta^\alpha \Gamma(\alpha)} \quad (\text{I.45})$$

$$F(x) = G\left(\alpha, \frac{x - \xi}{\beta}\right) / \Gamma(\alpha) \quad (\text{I.46})$$

Si $\gamma = 0$, la distribución es Normal, el rango de x es $-\infty < x < \infty$, y

$$f(x) = \phi\left(\frac{x - \mu}{\sigma}\right), \quad F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) \quad (\text{I.47})$$

Si $\gamma < 0$, el rango de x es $-\infty < x \leq \xi$, y

$$f(x) = \frac{(\xi - x)^{\alpha-1} e^{-(\xi-x)/\beta}}{\beta^\alpha \Gamma(\alpha)} \quad (\text{I.48})$$

$$F(x) = 1 - G\left(\alpha, \frac{\xi - x}{\beta}\right) / \Gamma(\alpha) \quad (\text{I.49})$$

En cada caso, $x(F)$ no tiene expresión analítica explícita. Γ es la función Gamma, definida en (I.23), y

$$G(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t} dt \quad (\text{I.50})$$

es la función Gamma incompleta; ϕ y Φ son la función de distribución Normal y su correspondiente acumulada.

Casos especiales: $\gamma = 2$ es la distribución exponencial; $\gamma = 0$ es la Normal; $\gamma = -2$ es la exponencial revertida.

La distribución de Pearson tipo III se remite generalmente al caso $\gamma > 0$, y se suele parametrizar por medio de α , β y ξ . La parametrización aquí expuesta se extiende a las distribuciones de Pearson tipo III más usuales. Esto permite usar la distribución de Pearson tipo III cuando los datos tienen una posible asimetría negativa. Los parámetros μ , σ y γ son los momentos convencionales de la distribución.

I.5.2.- L-MOMENTOS

Se muestran a continuación las expresiones de los L-momentos de la distribución referidos a sus parámetros estándar, por ser más sencillas. Se asume $\gamma > 0$. Los resultados en el caso $\gamma < 0$ serían los obtenidos al cambiar los signos de λ_1 , τ_3 y ξ en las expresiones (I.51) hasta (I.58).

Los L-momentos se definen para todos los valores de α entre 0 e ∞ .

$$\lambda_1 = \xi + \alpha\beta \quad (\text{I.51})$$

$$\lambda_2 = \pi^{-1/2} \beta \Gamma(\alpha + \frac{1}{2}) / \Gamma(\alpha) \quad (\text{I.52})$$

$$\tau_3 = 6I_{1/3}(\alpha, 2\alpha) - 3 \quad (\text{I.53})$$

donde

$$I_x(p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x t^{p-1} (1-t)^{q-1} dt \quad (\text{I.54})$$

No hay una expresión simple para τ_4 , si bien tanto τ_4 como τ_3 se pueden aproximar en función de α mediante funciones racionales. Las siguientes aproximaciones tienen una precisión del orden de 10^{-6} . Para $\alpha \geq 1$,

$$\tau_3 \approx \alpha^{-1/2} \frac{A_0 + A_1\alpha^{-1} + A_2\alpha^{-2} + A_3\alpha^{-3}}{1 + B_1\alpha^{-1} + B_2\alpha^{-2}} \quad (\text{I.55})$$

$$\tau_4 \approx \frac{C_0 + C_1\alpha^{-1} + C_2\alpha^{-2} + C_3\alpha^{-3}}{1 + D_1\alpha^{-1} + D_2\alpha^{-2}} \quad (\text{I.56})$$

Para $\alpha < 1$,

$$\tau_3 \approx \frac{1 + E_1\alpha + E_2\alpha^2 + E_3\alpha^3}{1 + F_1\alpha + F_2\alpha^2 + F_3\alpha^3} \quad (\text{I.57})$$

$$\tau_4 \approx \frac{1 + G_1\alpha + G_2\alpha^2 + G_3\alpha^3}{1 + H_1\alpha + H_2\alpha^2 + H_3\alpha^3} \quad (\text{I.58})$$

Los coeficientes de la aproximación son los recogidos en la tabla I.2.

I.5.3.- PARÁMETROS

Para estimar α se resuelve (I.53) reemplazando τ_3 por $|\tau_3|$ para posibilitar la resolución en caso de que τ_3 sea negativo. La siguiente aproximación tiene una precisión relativa mayor que 5×10^{-5} para cualquier valor de α . Para $0 < |\tau_3| < 1/3$ y definiendo $z = 3\pi \tau_3^2$,

$$\alpha \approx \frac{1 + 0.2906z}{z + 0.1882z^2 + 0.0442z^3} \quad (\text{I.59})$$

Para $1/3 \leq |\tau_3| < 1$ y con $z = 1 - |\tau_3|$,

$$\alpha \approx \frac{0.36067z + 0.59567z^2 + 0.25361z^3}{1 - 2.78861z + 2.56096z^2 + 0.77045z^3} \quad (\text{I.60})$$

Dado α , los parámetros definidos al principio de la sección se obtienen como sigue:

$$\gamma = 2\alpha^{-1/2} \text{sign}(\tau_3), \quad \sigma = \lambda_2 \pi^{1/2} \alpha^{1/2} \Gamma(\alpha) / \Gamma(\alpha + \frac{1}{2}), \quad \mu = \lambda_1 \quad (\text{I.61})$$

Tabla I.2. Coeficientes de las aproximaciones de τ_4 como τ_3 en función de α .

$A_0 = 3.2573501 \times 10^{-1}$	$C_0 = 1.2260172 \times 10^{-1}$
$A_1 = 1.6869150 \times 10^{-1}$	$C_1 = 5.3730130 \times 10^{-1}$
$A_2 = 7.8327243 \times 10^{-2}$	$C_2 = 4.3384378 \times 10^{-2}$
$A_3 = -2.9120539 \times 10^{-3}$	$C_3 = 1.1101277 \times 10^{-3}$
$B_1 = 4.6697102 \times 10^{-1}$	$D_1 = 1.8324466 \times 10^{-1}$
$B_2 = 2.4255406 \times 10^{-1}$	$D_2 = 2.0166036 \times 10^{-1}$
$E_1 = 2.3807576$	$G_1 = 2.1235833$
$E_2 = 1.5931792$	$G_2 = 4.1670213$
$E_3 = 1.1618371 \times 10^{-1}$	$G_3 = 3.1925299$
$F_1 = 5.1533299$	$H_1 = 9.0551443$
$F_2 = 7.1425260$	$H_2 = 2.6649995 \times 10^1$
$F_3 = 1.9745056$	$H_3 = 2.6193668 \times 10^1$

I.6.- DISTRIBUCIÓN KAPPA

I.6.1.- DEFINICIÓN

Parámetros (4): ξ (localización), α (escala), k , h .

Rango de x : el límite superior es $\xi + \alpha/k$ para $k > 0$, o bien ∞ para $k \leq 0$; el límite inferior es $\xi + \alpha(1 - h^{-k})/k$ para $h > 0$, $\xi + \alpha/k$ para $h \leq 0$ y $k < 0$, o bien $-\infty$ si $h \leq 0$ y $k \geq 0$.

$$f(x) = \alpha^{-1} [1 - k(x - \xi)/\alpha]^{1/k-1} [F(x)]^{1-h} \quad (\text{I.62})$$

$$F(x) = \left\{ 1 - h [1 - k(x - \xi)/\alpha]^{1/k} \right\}^{1/h} \quad (\text{I.63})$$

$$x(F) = \xi + \frac{\alpha}{k} \left[1 - \left(\frac{1 - F^h}{h} \right)^k \right] \quad (\text{I.64})$$

Casos especiales: $h = -1$ es la distribución Logística Generalizada; $h = 0$ es la distribución General de Valores Extremos; $h = 1$ es la distribución General de Pareto. La distribución Kappa de tres parámetros propuesta por Mielke y Johnson (1973) es un caso especial de la distribución Kappa definida aquí. Su función de distribución acumulada es

$$F(x) = (x/b)^\theta \left\{ a + (x/b)^{a\theta} \right\}^{-1/a}, \quad x \geq 0, \quad a, b, \theta > 0 \quad (\text{I.65})$$

y se obtiene de (I.63) tomando $\xi = b$, $\alpha = b/(a\theta)$, $k = -1/(a\theta)$ y $h = -a$.

La distribución Kappa es una distribución de cuatro parámetros que incluye como casos especiales la Logística Generalizada, la General de Valores Extremos y la General de Pareto. El rango de parámetros más útil es $h \geq -1$. Con esta restricción, los L-momentos de la distribución cubren un área amplia del plano (τ_3 , τ_4) (ver figura I.1). Es por ello que resulta de gran utilidad como distribución general con la cual comparar el grado de ajuste de distribuciones de dos o tres parámetros, y también a la hora de simular datos artificiales para evaluar la precisión de métodos estadísticos.

I.6.2.- L-MOMENTOS

Están definidos para $h \geq 0$ y $k > -1$, o bien para $h < 0$ y $-1 < k < -1/h$.

$$\lambda_1 = \xi + \alpha(1 - g_1)/k \quad (\text{I.66})$$

$$\lambda_2 = \alpha(g_1 - g_2)/k \quad (\text{I.67})$$

$$\tau_3 = (-g_1 + 3g_2 - 2g_3)/(g_1 - g_2) \quad (\text{I.68})$$

$$\tau_4 = (-g_1 + 6g_2 - 10g_3 + 5g_4)/(g_1 - g_2) \quad (\text{I.69})$$

donde

$$g_r = \begin{cases} \frac{r\Gamma(1+k)\Gamma(r/h)}{h^{1+k}\Gamma(1+k+r/h)}, & h > 0 \\ \frac{r\Gamma(1+k)\Gamma(-k-r/h)}{(-h)^{1+k}\Gamma(1-r/h)}, & h < 0 \end{cases} \quad (I.70)$$

El símbolo $\Gamma(\cdot)$ representa la función Gamma.

I.6.3.- PARÁMETROS

No hay expresiones simples para los parámetros en términos de L-momentos. Sin embargo, las expresiones (I.68) y (I.69) devuelven τ_3 y τ_4 a partir de k y h , de modo que dadas τ_3 y τ_4 se pueden obtener k y h por medio del método iterativo de Newton-Raphson. El algoritmo se describe en Hosking (1996).

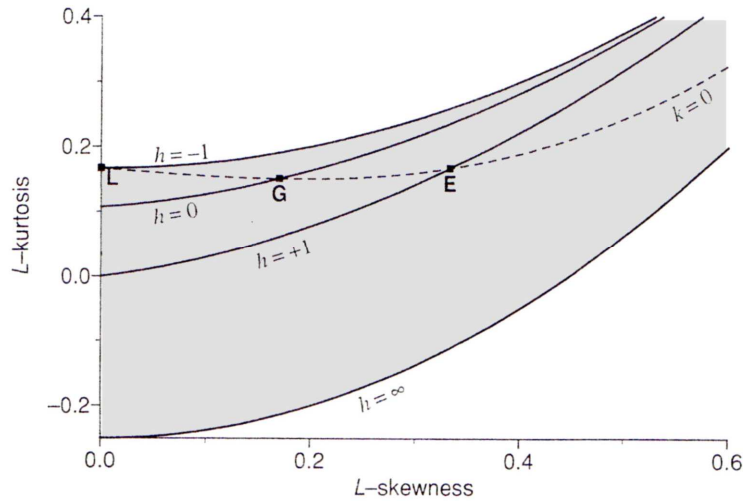


Fig. I.1. Diagrama del ratio de L-momentos para la distribución Kappa. El área sombreada muestra los valores de L-asimetría y L-kurtosis para $h \geq -1$.

I.7.- DISTRIBUCIÓN DE WAKEBY

I.7.1.- DEFINICIÓN

Parámetros (5): ξ (localización), α , β , γ , δ .

Rango de x : $\xi \leq x < \infty$ para $\delta \geq 0$ y $\gamma > 0$; $\xi \leq x \leq \xi + \alpha/\beta - \gamma/\delta$ para $\delta < 0$ y $\gamma = 0$.

$f(x)$, $F(x)$ no definidas explícitamente

$$x(F) = \xi + \frac{\alpha}{\beta} \left(1 - (1-F)^\beta \right) - \frac{\gamma}{\delta} \left(1 - (1-F)^{-\delta} \right) \quad (I.71)$$

Casos especiales: $\alpha = 0$ y $\gamma = 0$ dan como resultado la distribución de General de Pareto.

La parametrización aquí expuesta presenta la distribución de Wakeby como una generalización de la distribución General de Pareto, y proporciona estimaciones de α y γ estables ante pequeñas perturbaciones en los datos.

La distribución permanece invariante tras el cambio $\alpha \leftrightarrow \beta$ y $\gamma \leftrightarrow -\delta$, de modo que sin pérdida de generalidad se puede asumir que $\beta + \delta \geq 0$. De acuerdo con Hosking (1986b) y con Kotz, Johnson y Read (1988), se asume que se satisfacen también las siguientes condiciones (más fuertes):

- (i) $\beta + \delta > 0$ o bien $\beta = \gamma = \delta = 0$;
- (ii) si $\alpha = 0$, entonces $\beta = 0$;
- (iii) si $\gamma = 0$, entonces $\delta = 0$.

Para que $x(F)$ sea una función cuantil válida, se deben imponer además las condiciones

- (iv) $\gamma \geq 0$;
- (v) $\alpha + \delta \geq 0$.

La distribución de Wakeby ha sido utilizada en hidrología para modelar datos de caudales. Las siguientes propiedades son las que hacen de esta una distribución particularmente interesante para aplicaciones en ciencias ambientales:

- para valores adecuados de los parámetros, la distribución de Wakeby puede mimetizar las formas de diversas distribuciones asimétricas conocidas, tales como la de Valores Extremos, la LogNormal o la de Pearson tipo III;
- tiene 5 parámetros, más que muchas otras distribuciones de uso común, lo que le permite adoptar más formas (esto la hace especialmente útil para simular datos artificiales en estudios sobre la robustez de métodos de análisis de datos respecto a cambios en la forma de la distribución);
- cuando $\delta > 0$, tiene una cola superior pronunciada y puede por tanto ajustarse bien a datos que presentan ocasionalmente outliers, fenómeno que se observa frecuentemente en ciencias ambientales;
- tiene un límite inferior finito, lo cual es físicamente razonable para observaciones en el mundo real;
- la forma explícita de la función cuantil, expresión (I.71), facilita simular muestras aleatorias a partir de la distribución de Wakeby.

I.7.2.- L-MOMENTOS

Los L-momentos se definen para $\delta < 1$.

$$\lambda_1 = \xi + \frac{\alpha}{1+\beta} + \frac{\gamma}{1-\delta} \quad (\text{I.72})$$

$$\lambda_2 = \frac{\alpha}{(1+\beta)(2+\beta)} + \frac{\gamma}{(1-\delta)(2-\delta)} \quad (\text{I.73})$$

$$\lambda_3 = \frac{\alpha(1-\beta)}{(1+\beta)(2+\beta)(3+\beta)} + \frac{\gamma(1+\delta)}{(1-\delta)(2-\delta)(3-\delta)} \quad (\text{I.74})$$

$$\lambda_4 = \frac{\alpha(1-\beta)(2-\beta)}{(1+\beta)(2+\beta)(3+\beta)(4+\beta)} + \frac{\gamma(1+\delta)(2+\delta)}{(1-\delta)(2-\delta)(3-\delta)(4-\delta)} \quad (\text{I.75})$$

No hay una expresión simple para τ_r .

I.7.3.- PARÁMETROS

En términos de L-momentos, el algoritmo de Hosking (1996) proporciona las expresiones siguientes para los parámetros. Para ξ desconocido,

$$\begin{aligned} N_1 &= 3\lambda_2 - 25\lambda_3 + 32\lambda_4 & C_1 &= 7\lambda_2 - 85\lambda_3 + 203\lambda_4 - 125\lambda_5 \\ N_2 &= -3\lambda_2 + 5\lambda_3 + 8\lambda_4 & C_2 &= -7\lambda_2 + 25\lambda_3 + 7\lambda_4 - 25\lambda_5 \\ N_3 &= 3\lambda_2 + 5\lambda_3 + 2\lambda_4 & C_3 &= 7\lambda_2 + 5\lambda_3 - 7\lambda_4 - 5\lambda_5 \end{aligned} \quad (\text{I.76})$$

Así, β y $-\delta$ son las raíces de la ecuación cuadrática

$$(N_2C_3 - N_3C_2)z^2 + (N_1C_3 - N_3C_1)z + (N_1C_2 - N_2C_1) = 0 \quad (\text{I.77})$$

siendo β la mayor de las dos, mientras que el resto de parámetros vienen dados por

$$\alpha = (1+\beta)(2+\beta)(3+\beta)[(1+\delta)\lambda_2 - (3-\delta)\lambda_3]/[4(\beta+\delta)] \quad (\text{I.78})$$

$$\gamma = -(1-\delta)(2-\delta)(3-\delta)[(1-\beta)\lambda_2 - (3+\beta)\lambda_3]/[4(\beta+\delta)] \quad (\text{I.79})$$

$$\xi = \lambda_1 - \alpha/(1+\beta) - \gamma/(1-\delta) \quad (\text{I.80})$$

Para ξ conocido, asumiendo sin perder generalidad que $\xi = 0$,

$$\begin{aligned} N_1 &= 4\lambda_1 - 11\lambda_2 + 9\lambda_3 & C_1 &= 10\lambda_1 - 29\lambda_2 + 35\lambda_3 - 16\lambda_4 \\ N_2 &= -\lambda_2 + 3\lambda_3 & C_2 &= -\lambda_2 + 5\lambda_3 - 4\lambda_4 \\ N_3 &= \lambda_2 + \lambda_3 & C_3 &= \lambda_2 - \lambda_4 \end{aligned} \quad (\text{I.81})$$

Nuevamente, β y $-\delta$ son las raíces de la ecuación (I.77), siendo β la mayor de ellas. En este caso, el resto de parámetros se calculan como

$$\alpha = (1+\beta)(2+\beta)[\lambda_1 - (2-\delta)\lambda_2]/(\beta+\delta) \quad (\text{I.82})$$

$$\gamma = -(1-\delta)(2-\delta)[\lambda_1 - (2+\beta)\lambda_2]/(\beta+\delta) \quad (\text{I.83})$$

ANEXO II: IMPLEMENTACIÓN EN R

Antes de comenzar con el análisis regional, es necesario descargar e instalar los paquetes *lmom* (Hosking, 2009) y *lmomRFA* (Hosking, 2010) en la configuración básica de R.

Una vez instalados, se cargan mediante la siguiente orden:

```
> library(lmom)
> library(lmomRFA)
```

Para facilitar el estudio, se asocia a cada estación un vector a su nombre con sus valores máximos anuales de precipitación en 24 horas filtrados y depurados:

```
> belzunce<-c
(60,53.2,61,84,122,57,38,76.7,80.5,35.5,44,63,31,48,45,41,72,63,
56,40)

>erro<-c
(76.2,46,43.3,46.2,73.3,67.5,69.5,57,57.5,83.5,60,49,41.3,60,100,
88,56,70,110,44,56,61.3,71.5,62.5,43,86.8,33.5,55.8,60.5,47,77,
70.7,67.4,50.7)

>espinal<-c
(50.2,90,56.5,49.9,54.9,48,90,123,68,65.5,76.5,58.5,94,47.5,49.8,
30,32.5,46,49.1,33,92.4,45.7,34.8,29.8,72.4,65.5,58.2,51.3,88.4,
47.4,70.2,52.1,63.8,93.5,85.1,73.4,68.5)

>eugui<-c
(51.8,90,45.5,76,76.3,63.9,63.3,109.4,65.5,49.5,56.7,110,95.5,81,
74.5,78.8,52,64.2,48,52.5,58.6,58,56,92.6,114,57.3,51.4,58.2,
81,63.2,53.5,75.5,34,62.5,64,52.5,65.4,92.4,61.6,64.5)

>ilundain<-c
(51.5,41,72,63,36,35,36,36,76.5,51,39.7,50.3,65.6,43.5,41,65.6,6,
1.5,35.6,80.5,68.7,36,46,50,35.4,49.8,51,45.3,32)

>iraizoz<-c
(46.3,59.9,85,70.2,63,82,91.5,55,52,59,45,55,50,50,63,65,100,45,
53.3,55,75,51.3,52,45.9,60.9,87.5,60.1,52.4,72.5,78.4,50.4)

>iroz<-c
(60,40,36.5,57,40,30,29.5,42,84,50,22,36,50,37,50,60,38,42.5,59,
23,38,50,55,69,60,53.6,40)

>monreal<-c
(98.3,39.3,64.5,29.8,38,36,31.3,71.5,21,35,53.3,65.5,45,41.5,56.
2,42.6,33.9,93,57.5,37.5,34.9,57.6,38.5,47,66.5,50,45)

>noain<-c
(96.5,54.5,29.7,49.2,107.4,39.6,31.9,33,64.6,42.3,38.2,34.2,41.3,
47,86.7,25.6,34.6,30.4,53.8,28.6,30.5,54.4,51.2,37.8,50.3,61.3,
20.7,33.7,42.8,27.1,47.2,61,38.4,45.1)

> olagüe<-c
(82,65,51,59,53,45,60.5,63,56.5,47.5,40,62.5,53,60,56.5,90.8,86.
6,41.5,38,41.5,83,38,46,69,30,49,88,42,67.5,58,52,42)
```

```
> otazu<-c
(41.5,27.5,37.5,56,39,48,28.5,38,45.5,43.2,35.5,56.4,66.5,66.4,3
4.7,44.5,59,79,35.5,44,65.7,73.7,33,42,38.1,39.5,30.5,39.5,38.5,
56.3,50.5,57,57.8)

>pamplona<-c
(26.8,40.4,49.2,46.5,34.8,79.3,30,50,56.1,28.5,40,58.3,39,46.2,3
4.5,33.9,37.5,31.8,42.7,41.3,73.7,31.3,36.8,42.7,63,86,48,24,34.
8,38.4,77,58,53.5,39.4,32.5,33.5,32.9,35,34.1,72,40,85,96,80,45.
2,65.6,46,135,78.4,101.6,82.4,107,65.2,82,112,78,33,46,40.2,32.4
,40,21,79.1,25,43.6,46,46.1,85.2,64,91.2,115.5,50.4,40.4,54.6,53
.7,57.9,50.8,64.2,88.8,43.2,48.1,47.2,39.8,67.3,79,46.6,31.8,65.
4,41,32.4,31.2,42.6,43.8,40.5,25.1,35.8,39.1,90.8,43.6,45.1,41.6
,69.2,36.5,35.4,43.5,63,57.4,62.1,21.7,41.9,55.2,43.3,63.3,67.6,
51.5,48.7)

>velate<-c
(52.5,78.4,61.5,46.6,81.6,81,83.1,85.3,95,84,57.4,45.2,66.5,53,6
7.1,54.2,50.5,70.4,78.4,67,50)

>zuazu<-c
(45.6,44.3,30.7,51.3,39.2,59.5,48.7,43.9,51.3,64.2,28.3,87.5,66.
6,36.4,31.6,37.8,32.5,57,106.7,53.3,42)

> zubiri<-c
(71.7,47.2,50.6,69.4,60,54.5,42,35,56,54,100,78.5,58.5,59,35,58,
63,56,56,60,82.5,43,38,50,67.5,53,42,74.5,33,40,57,51,54.9,73,53
.5,42.3)
```

La construcción de la tabla 3.1 con los ratios de L -momentos y las longitudes muestrales de las estaciones se lleva a cabo mediante el comando *regsamlmu*, creando previamente una lista de todas ellas. Se nombra *A* al objeto creado, y con el argumento *nmom* se especifica el orden de ratios de L -momentos que interesa calcular:

```
> estaciones<-list
(belzunce=belzunce,erro=erro,espinal=espinal,eugui=eugui,ilundai
n=ilundain,iraizoz=iraizoz,iroz=iroz,monreal=monreal,noain=noain
,olagüe=olagüe,otazu=otazu,pamplona=pamplona,velate=velate,zuazu
=zuazu,zubiri=zubiri)
> A<-regsamlmu(estaciones,nmom=5)
> A
```

1. FILTRADO PRIMARIO DE LAS OBSERVACIONES

La creación de los gráficos de la figura 3.1 para la primera inspección visual se implementa mediante la orden *plot*, configurando nombres de ejes, rangos de dichos ejes, tamaños de letra, valores a representar, y estaciones a identificar manualmente:

```
> par(mfrow=c(1,2),pty="s")

> plot(A$t_3, A$t,xlim=c(0,0.4), ylim=c(0,0.4),
xlab=expression(paste("L-asimetría", italic(t[3]))),
ylab=expression(paste("L-CV", italic(t))), cex.lab=0.9,
cex.axis=0.8, col="red", pch=20)
```

```

> points(mean(A$t_3),mean(A$t),pch=3,lwd=1.5)
> identify(A$t_3,A$t,labels=A$name,cex=0.7)

> plot(A$t_3, A$t_4, xlim=c(0,0.4), ylim=c(-0.1,0.4),
xlab=expression(paste("L-asimetría", italic(t[3]))),
ylab=expression(paste("L-kurtosis", italic(t[4]))),
cex.lab=0.9, cex.axis=0.8, col="blue", pch=20)
> points(mean(A$t_3),mean(A$t_4),pch=3,lwd=1.5)
> identify(A$t_3,A$t_4,labels=A$name,cex=0.7)

```

El cálculo de la medida de la discordancia para obtener la tabla 3.2 se ejecuta mediante el comando *regtst* aplicado a la tabla de ratios y longitudes muestrales (objeto A creado):

```

> set.seed(123)
> regtst(A)

```

2. IDENTIFICACIÓN DE REGIONES HOMOGÉNEAS

Para la creación de la tabla 3.3 de los descriptores físicos de las estaciones, se vectoriza una a una cada una de las variables antes de ejecutar el cluster:

```

> estacion<-c
("belzunce","erro","espinal","eugui","ilundain","iraizoz","iroz",
,"monreal","noain","olagüe","otazu","pamplona","velate","zuazu",
,"zubiri")

> latitud_N<-c
(42.898297,42.942428,42.978861,42.978589,42.775703,42.993322,42.
861214,42.704636,42.760461,42.963483,42.784906,42.817992,43.0490
47,42.736708,42.930431)

> altitud<-c
(530,671,870,616,572,550,478,533,456,549,387,453,797,627,526)

> T_med_an<-c
(11.7,9.8,8.8,10.3,12.1,10.8,12.4,13.2,12.4,11.1,12.5,12.5,10.5,
12.6,11.5)

> P_med_an<-c
(1100,1310.3,1521.1,1535.8,837.1,1321.8,890.8,783.1,714.8,1246.5
,807.1,813.4,1450,825.9,1225)

```

Los vectores se compilan en una tabla (a la que se nombra B) mediante el comando *data.frame*, y ésta, para poder ser operada, se transforma en una matriz numérica (a la que se nombra “características”) mediante el comando *c.bind*:

```

> B<-data.frame(estacion,latitud_N,altitud,T_med_an,P_med_an)
> B

> caracteristicas<-cbind
(a1=B$latitud_N,a2=B$altitud,a3=B$T_med_an,a4=P_med_an)

```

```
> características
```

La estandarización y ponderación de cada variable se ejecuta por columnas sobre la matriz “características” mediante el comando *apply*:

```
> características<-apply(características,2,function(x) x/sd(x))

> características[,2]<-características[,2]*1.5
> características[,3]<-características[,3]*0.5
> características[,4]<-características[,4]*1.5
```

Por último, se ejecuta el cluster mediante los comandos *cluagg* y *cluinf* aplicados en la matriz “características” modificada. En el argumento *method* que acompaña a *cluagg* se escoge el algoritmo ward, mientras que en *cluinf* se especifica el número de clusters resultantes deseado:

```
> cl<-cluagg(características,method="ward")
> cluinf(cl,2)
```

La definición de los clusters resultantes se ejecuta procediendo del mismo modo que en la construcción de la tabla inicial A, mediante el comando *regsamlmu* aplicado a la lista de las estaciones de cada región previamente definida. Así, se crean las regiones Norte (nombrada D) y Sur (nombrada C):

```
# Región Norte:
>norte<-list
(belzunce=belzunce,erro=erro,espinal=espinal,eugui=eugui,iraizoz
=iraizoz,iroz=iroz,olagüe=olagüe,velate=velate,zubiri=zubiri)
> D<-regsamlmu(norte,nmom=5)
> D

# Región Sur:
> sur<-list
(ilundain=ilundain,monreal=monreal,noain=noain,otazu=otazu,pampl
ona=pamplona,zuazu=zuazu)
> C<-regsamlmu(sur,nmom=5)
> C
```

El cálculo de las medidas de heterogeneidad y discordancia para cada región resultante (tablas 3.6 y 3.7) se implementa también con el comando *regtst*, incluyendo en esta ocasión el argumento *nsim* para especificar el número de simulaciones a realizar:

```
# Región Norte:
> set.seed(123)
> regtst(D,nsim=500)

# Región Sur:
> set.seed(123)
> regtst(C,nsim=500)
```

3. ELECCIÓN DE LA DISTRIBUCIÓN DE FRECUENCIA

Previamente a la elección de la distribución, se introducen las matrices de correlaciones (tablas 3.8 y 3.9) de la región Norte (nombrada D) y Sur (nombrada C), calculadas por comparación de cada par de estaciones mediante el comando *cor*:

```
# Región Norte:
> corD<-matrix
(c(1,0.6894845,0.2328556,0.8134707,0.5698033,0.2735637,0.6904301,
,0.5821075,0.7356203,0.6894845,1,0.2671608,0.6212587,0.3272598,0.
.6814821,0.5806987,0.4152367,0.6383243,0.2328556,0.2671608,1,0.4
549886,0.2560522,0.4116879,0.2050988,0.2116244,0.2136551,0.81347
07,0.6212587,0.4549886,1,0.6924327,0.3375959,0.609299,0.4493284,
0.665939,0.5698033,0.3272598,0.2560522,0.6924327,1,0.1676848,0.6
049622,0.1510664,0.368918,0.2735637,0.6814821,0.4116879,0.337595
9,0.1676848,1,0.3345877,0.4139183,0.5296346,0.6904301,0.5806987,
0.2050988,0.609299,0.6049622,0.3345877,1,0.2087957,0.5601548,0.5
821075,0.4152367,0.2116244,0.4493284,0.1510664,0.4139183,0.20879
57,1,0.2907645,0.7356203,0.6383243,0.2136551,0.665939,0.368918,0
.5296346,0.5601548,0.2907645,1),nrow=9,ncol=9,byrow=TRUE,dimname
s=list(c("belzunce","erro","espinal","eugui","iraizoz","iroz","o
lagüe","velate","zubiri"),c("belzunce","erro","espinal","eugui",
"iraizoz","iroz","olagüe","velate","zubiri"))

# Región Sur:
> corC<-matrix
(c(1,0.7938805,0.7078224,0.2797085,0.6441765,0.5046491,0.7938805
,1,0.6221021,0.2814581,0.4473492,0.6680827,0.7078224,0.6221021,1
,0.4302550,0.7520877,0.5052279,0.2797085,0.2814581,0.4302550,1,0
.5203773,0.2619512,0.6441765,0.4473492,0.7520877,0.5203773,1,0.4
742516,0.5046491,0.6680827,0.5052279,0.2619512,0.4742516,1),nrow
=6,ncol=6,byrow=TRUE,dimnames=list(c("ilundain","monreal","noain
","otazu","pamplona","zuazu"),c("ilundain","monreal","noain","ot
azu","pamplona","zuazu"))
```

Se programa para cada región su región simulada, definiéndola con igual número de estaciones y longitudes muestrales, pero con ratios de *L*-momentos escogidos en un rango arbitrariamente inferior al observado. Se ajusta a la distribución Kappa:

```
# Región Norte:
> rmomD<-regavlmom(D)
> nsitesD<-nrow(D)
> meansD<-rep(1,nsitesD)
> LCVsD<-c
(0.1852298,0.1587298,0.1786048,0.1521048,0.1388548,0.1719798,0.1
653548,0.1322298,0.1454798)
> LskewsD<-c
(0.21230667,0.15605667,0.13730667,0.19355667,0.23105667,0.099806
67,0.17480667,0.08105667,0.11855667)
> LkurtD<-c
(0.17975462,0.13850462,0.09725462,0.15225462,0.08350462,0.124754
62,0.11100462,0.06975462,0.16600462)
> ppD<-t(apply(cbind(meansD,meansD*LCVsD,LskewsD,LkurtD), 1,
pelkap))
```

```
# Región Sur:
> rmomC<-regavlmom(C)
> nsitesC<-nrow(C)
> meansC<-rep(1,nsitesC)
> LCVsC<-c
(0.1787476,0.2087476,0.2187476,0.1887476,0.2287476,0.1987476)
> LskewsC<-c
(0.2212723,0.2392723,0.2932723,0.2032723,0.2572723,0.2752723)
> LkurtC<-c
(0.06252562,0.15852562,0.19052562,0.09452562,0.12652562,0.222525
62)
> ppC<-t(apply(cbind(meansC,meansC*LCVsC,LskewsC,LkurtC), 1,
pelkap))
```

Finalmente, se utiliza el comando *regsimh* en lugar de *regtst* para considerar los valores de correlación y computar, en base a las regiones simuladas programadas, las medidas definitivas de heterogeneidad y bondad de ajuste (tablas 3.10, 3.11 y 3.13):

```
# Región Norte:
> set.seed(123)
>
regsimh(qfunc=quakap,para=ppD,cor=corD,nrec=D$n,nrep=1000,nsim=1
00)

# Región Sur:
> set.seed(123)
>
regsimh(qfunc=quakap,para=ppC,cor=corC,nrec=C$n,nrep=1000,nsim=1
00)
```

4. ESTIMACIÓN DE LOS CUANTILES DE FRECUENCIA

Para el ajuste a una distribución concreta (tablas 3.15 y 3.28), se emplea el comando *regfit* acompañado de dos argumentos: el objeto de datos a los que se quiere ajustar (D para la región Norte y C para la región Sur), y las iniciales de la distribución de interés (“glo”, “gev”, “gno”, “pe3” o “gpa”):

```
# Región Norte:
> ajusteD_PE3<-regfit(D,"pe3")
> ajusteD_PE3

# Región Sur:
> ajusteC_PE3<-regfit(C,"pe3")
> ajusteC_PE3
```

Definido el ajuste, los comandos *regquant* y *sitequant* permiten calcular cuantiles - regionales y locales, respectivamente – especificados mediante un vector de probabilidades de no excedencia:

```
# Región Norte:
> regquant
(c(0.5,0.8,0.9,0.95,0.98,0.99,0.995,0.998, 0.999),ajusteD_PE3)
```



```
# Para la estación de Eugui (sitio 4 de la lista D):
> sitequant
(c(0.5,0.8,0.9,0.95,0.98,0.99,0.995,0.998,
0.999),ajusteD_PE3,sitenames=4)

# Región Sur:
> regquant
(c(0.5,0.8,0.9,0.95,0.98,0.99,0.995,0.998, 0.999),ajusteC_PE3)
# Para la estación de Pamplona (sitio 5 de la lista C):
> sitequant
(c(0.5,0.8,0.9,0.95,0.98,0.99,0.995,0.998,
0.999),ajusteC_PE3,sitenames=5)
```

Para establecer los límites de error al 90% y el ECM de las estimaciones, se programa para cada región su región simulada, definiéndola con igual número de estaciones y longitudes muestrales, pero con ratios de *L*-momentos escogidos en un rango arbitrariamente inferior al observado. Se emplea el comando *regsimq*. Evidentemente, en esta ocasión se ajusta a la distribución elegida:

```
# Región Norte:
> rmomD<-regavlmom(D)
> nsitesD<-nrow(D)
> meansD<-(D$l_1)
> LCVsD<-c
(0.1852298,0.1587298,0.1786048,0.1521048,0.1388548,0.1719798,0.1
653548,0.1322298,0.1454798)
> LskewsD<-c
(0.21230667,0.15605667,0.13730667,0.19355667,0.23105667,0.099806
67,0.17480667,0.08105667,0.11855667)
> ppD_PE3<-t(apply(cbind(meansD,meansD*LCVsD,LskewsD),1,pelpe3))
> corD<-matrix
(c(1,0.6894845,0.2328556,0.8134707,0.5698033,0.2735637,0.6904301
,0.5821075,0.7356203,0.6894845,1,0.2671608,0.6212587,0.3272598,0
.6814821,0.5806987,0.4152367,0.6383243,0.2328556,0.2671608,1,0.4
549886,0.2560522,0.4116879,0.2050988,0.2116244,0.2136551,0.81347
07,0.6212587,0.4549886,1,0.6924327,0.3375959,0.609299,0.4493284,
0.665939,0.5698033,0.3272598,0.2560522,0.6924327,1,0.1676848,0.6
049622,0.1510664,0.368918,0.2735637,0.6814821,0.4116879,0.337595
9,0.1676848,1,0.3345877,0.4139183,0.5296346,0.6904301,0.5806987,
0.2050988,0.609299,0.6049622,0.3345877,1,0.2087957,0.5601548,0.5
821075,0.4152367,0.2116244,0.4493284,0.1510664,0.4139183,0.20879
57,1,0.2907645,0.7356203,0.6383243,0.2136551,0.665939,0.368918,0
.5296346,0.5601548,0.2907645,1),nrow=9,ncol=9,byrow=TRUE,dimname
s=list(c("belzunce","erro","espinal","eugui","iraizoz","iroz","o
lagüe","velate","zubiri"),c("belzunce","erro","espinal","eugui",
"iraizoz","iroz","olagüe","velate","zubiri")))
> set.seed(123)
> sim_ajusteD_PE3<-regsimq
(qfunc=quape3,para=ppD_PE3,cor=corD,index=D$l_1,nrec=D$n,nrep=10
0,fit="pe3",f=c(0.5,0.8,0.9,0.95,0.98,0.99,0.995,0.998, 0.999))

# Región Sur:
> rmomC<-regavlmom(C)
> nsitesC<-nrow(C)
> meansC<-(C$l_1)
> LCVsC<-c
(0.1787476,0.2087476,0.2187476,0.1887476,0.2287476,0.1987476)
```

```
> LskewsC<-c
(0.2212723,0.2392723,0.2932723,0.2032723,0.2572723,0.2752723)
> ppC_PE3<-t(apply(cbind(meansC,meansC*LCVsC,LskewsC),1,pelpe3))
> corC<-matrix
(c(1,0.7938805,0.7078224,0.2797085,0.6441765,0.5046491,0.7938805
,1,0.6221021,0.2814581,0.4473492,0.6680827,0.7078224,0.6221021,1
,0.4302550,0.7520877,0.5052279,0.2797085,0.2814581,0.4302550,1,0
.5203773,0.2619512,0.6441765,0.4473492,0.7520877,0.5203773,1,0.4
742516,0.5046491,0.6680827,0.5052279,0.2619512,0.4742516,1),nrow
=6,ncol=6,byrow=TRUE,dimnames=list(c("ilundain","monreal","noain
","otazu","pamplona","zuazu"),c("ilundain","monreal","noain","ot
azu","pamplona","zuazu"))))
> set.seed(123)
> sim_ajusteC_PE3<-regsimq
(qfunc=quape3,para=ppC_PE3,cor=corC,index=C$l_1,nrec=C$n,nrep=10
0,fit="pe3",f=c(0.5,0.8,0.9,0.95,0.98,0.99,0.995,0.998, 0.999))
```

Los comandos *regquantbounds* y *sitequantbounds* permiten computar los errores a nivel regional y local, respectivamente:

```
# Región Norte:
> regquantbounds(sim_ajusteD_PE3,ajusteD_PE3)
# Para la estación de Eugui (sitio 4 de la lista D):
> sitequantbounds(sim_ajusteD_PE3,ajusteD_PE3,sitenames=4)

# Región Sur:
> regquantbounds(sim_ajusteC_PE3,ajusteC_PE3)
# Para la estación de Pamplona (sitio 5 de la lista C):
> sitequantbounds(sim_ajusteC_PE3,ajusteC_PE3,sitenames=5)
```

Por último, la orden *evplot* dibuja las curvas de ajuste tanto a nivel regional como local:

```
# Región Norte
> set.seed(123)
> sim_ajusteD_PE3bis<-regsimq
(qfunc=quape3,para=ppD_PE3,cor=corD,index=D$l_1,nrec=D$n,nrep=10
0,fit="pe3")
# Para la región:
> evplot
(ajusteD_PE3, regquantbounds (sim_ajusteD_PE3bis,ajusteD_PE3),
xlab=expression ("Variable Reducida Gumbel, " * -log(-
log(italic(F)))), ylab="Cuantil regional (adimensional)",
xlim=c(0,6.5),col="blue")
# Para la estación de Eugui (sitio 4 de la lista D):
> evplot
(ajusteD_PE3, sitequantbounds
(sim_ajusteD_PE3bis,ajusteD_PE3,sitenames=4), xlab=expression
("Variable Reducida Gumbel, " * -log(-log(italic(F)))),
ylab=expression(paste("Cuantil local ", P[24][h]," (mm)")),
xlim=c(0,6.5),col="blue")

# Región Sur
> set.seed(123)
> sim_ajusteC_PE3bis<-regsimq
```

```

(qfunc=quape3,para=ppC_PE3,cor=corC,index=C$l_1,nrec=C$n,nrep=10
0,fit="pe3")
# Para la región:
> evplot
(ajusteC_PE3, regquantbounds (sim_ajusteC_PE3bis,ajusteC_PE3),
xlab=expression ("Variable Reducida Gumbel, " * -log(-
log(italic(F)))), ylab="Cuantil regional (adimensional)",
xlim=c(0,6.5),col="blue")
# Para la estación de Pamplona (sitio 5 de la lista C):
> evplot
(ajusteC_PE3, sitequantbounds
(sim_ajusteC_PE3bis,ajusteC_PE3,sitenames=5), xlab=expression
("Variable Reducida Gumbel, " * -log(-log(italic(F)))),
ylab=expression(paste("Cuantil local ", P[24][h]," (mm)")),
xlim=c(0,6.5),col="blue")

```

